

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 9, 2003, 15:49:18 ; Search time 48 Seconds  
(without alignments)  
2195.718 Million cell updates/sec

Title: US-09-590-991-6  
Perfect score: 4398  
Sequence: 1 YPVVLAIDTSSSEDALNISDK.....PGKNIKSTLVINGKSTGY 664

Scoring table: BIOSUM30  
Gapop 8.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: A\_Geneseq\_19Jun03.\*  
2: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.\*  
23: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.\*  
24: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.\*  
25: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4398	100.0	664	22	AAB48342
2	4398	100.0	2120	21	AA1981.710
3	4380.5	99.6	2140	24	ABU01020
4	4171	94.8	1007	19	AAW61246
5	4171	94.8	1007	23	ABP54664
6	975.5	22.2	192	19	AA186159
7	962.5	21.9	1962	12	AA10559
8	954.5	21.7	1968	12	AA10941
9	953.5	21.7	1962	12	AA10557

10	953.5	21.7	1974	12	AA10940
11	951.5	21.6	1962	12	AA10558
12	948.5	21.6	1962	12	AA10560
13	948.5	21.6	1962	12	AA10561
14	948.5	21.6	1962	12	AA10563
15	947	21.5	1959	12	AA10562
16	938.5	21.3	1560	10	AA10562
17	896.5	20.4	1946	19	AAW47273
18	875	19.9	1570	23	ABP29894
19	875	19.9	1570	23	ABP29894
20	875	19.9	1570	23	ABP29894
21	873	19.8	1550	23	ABP29894
22	869.5	19.8	885	23	ABP27332
23	868	19.7	885	23	ABP29796
24	867.5	19.7	1239	23	ABP25832
25	865.5	19.7	1579	24	ABP71283
26	864	19.6	1233	23	ABP30203
27	862	19.6	1233	23	ABP29675
28	857	19.5	1647	23	ABP28561
29	797.5	18.1	690	11	AA104904
30	792	18.0	2710	17	AA105016
31	792	18.0	2710	19	AAW68387
32	787.5	17.9	690	18	AAW16337
33	787.5	17.9	690	18	AAW1787
34	782	17.8	806	13	AA127481
35	780.5	17.7	2835	23	AB198574
36	774	17.6	815	23	AB198574
37	771	17.5	922	19	AAW37372
38	769	17.5	1475	11	AA108221
39	764.5	17.4	1582	23	AB147307
40	761.5	17.3	903	17	AA107007
41	761.5	17.3	1398	17	AA107008
42	761.5	17.3	1398	18	AAW42124
43	761.5	17.3	1398	20	AAW4839
44	761	17.3	1941	23	ABP5509
45	760.5	17.3	1185	13	AA122675

#### ALIGNMENTS

RESULT 1  
ID AAB48342 standard; Protein: 664 AA.  
AC AAB48342;  
DT 20-APR-2001 (first entry)  
XX S. pneumoniae Sp128 polypeptide.  
XX Immunogenic; Sp128; Sp130; pneumococcal; otitis media; nasopharyngeal;  
XX bronchial; lung; blood; infection; immune response; immunotherapy;  
XX antibacterial; auditory; vaccine.  
XX Streptococcus pneumoniae.  
XX WO200076540-A2.  
XX 21-DEC-2000.  
XX 09-JUN-2000; 2000WO-US15925.  
XX 10-JUN-1999; 99US-0138453.  
XX (MEDI-) MED IMMUNE INC.  
XX Adamou JE, Choi GH;  
XX WPI; 2001-112197/12.  
XX N-PSDB; AAC84741.  
XX New vaccines comprising Sp128 or Sp130 polypeptides, for treating and



```

QY 181 GVEEAIDYLKSIINAFPGNFDGRGVISNIDTGTDIRRHKARIDDDAKASRPFKEDLKG 240
DB 181 GVEEAIDYLKSIINAFPGNFDGRGVISNIDTGTDIRRHKARIDDDAKASRPFKEDLKG 240
QY 241 TDKKYYWLSDKIPHAFFNYNGSKITVEKYDDGRDYFDPHGMIAGIAGLNDTEODIKPNFG 300
DB 241 TDKKYYWLSDKIPHAFFNYNGSKITVEKYDDGRDYFDPHGMIAGIAGLNDTEODIKPNFG 300
QY 301 IDGIAPNAQISYKMYSDAGSGFAGDETFMFAIEDSIKHNVDVSVSSGFTGTGLVGEKY 360
DB 301 IDGIAPNAQISYKMYSDAGSGFAGDETFMFAIEDSIKHNVDVSVSSGFTGTGLVGEKY 360
QY 361 WQAIRALRKAGIPMVVATGNVATSASSSSMDLVANNHLMKMTDGTGNVTRTAHEDAIAVAS 420
DB 361 WQAIRALRKAGIPMVVATGNVATSASSSSMDLVANNHLMKMTDGTGNVTRTAHEDAIAVAS 420
QY 421 AKNOTVEFDKVNIGESFKYRNIGAFPDKSKITTNEDGTAPSKLKFYIGKGODDLIG 480
DB 421 AKNOTVEFDKVNIGESFKYRNIGAFPDKSKITTNEDGTAPSKLKFYIGKGODDLIG 480
QY 481 LDLRGKIAVMDRITTKDLKNAFKKAMDGGARAIVVNVTVNVRNDMTLPMGVEADEG 540
DB 481 LDLRGKIAVMDRITTKDLKNAFKKAMDGGARAIVVNVTVNVRNDMTLPMGVEADEG 540
QY 541 TKSQVFSISGDDGVKLMNMINPDKTEYKRNKEDFKDLKQYYPIDMESFNSNKPNGD 600
DB 541 TKSQVFSISGDDGVKLMNMINPDKTEYKRNKEDFKDLKQYYPIDMESFNSNKPNGD 600
QY 601 EKEIDFKFAPDTDELYKEDIIVPAGSTSWGPRIDLLKPVSAFGKNIKSTLVNINGS 660
DB 601 EKEIDFKFAPDTDELYKEDIIVPAGSTSWGPRIDLLKPVSAFGKNIKSTLVNINGS 660
QY 661 TYGY 664
DB 661 TYGY 664

```

RESULT 3  
ABU01020 standard; Protein; 2140 AA.

```

XX AC ABU01020;
XX DT 11-FEB-2003 (first entry)
XX DE S. pneumoniae type 4 strain protein from coding region #590.
XX KW Bacterial meningitis; pneumonia; sepsis; otitis media;
XX KW ear infection; antiinflammatory; antibacterial; immunostimulant;
XX KW auditory; respiratory; gene therapy; vaccine.
XX OS Streptococcus pneumoniae type 4 strain.
XX PN MO200277021-A2.
XX PD 03-OCT-2002.
XX PF 27-MAR-2002; 2002WO-IB02163.
XX PR 27-MAR-2001; 2001GB-0007658.
XX PA (CHIR-) CHIRON SPA.
XX PA (GENO-) INST GENOMIC RES.
XX PI Maignant V, Tectelin H, Fraser C;
XX DR MPI; 2003-040579/03.
XX DR N-PSDB; ABX06302.
XX PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,
PT useful as medicaments for treating or preventing a disease or infection
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media
PT or ear infection

```

```

XX XX Claim 1; SEQ ID No 1180; 56bp; English.
PS XX
CC The invention relates to a protein comprising or having at least 50%
CC identity to any of the 2469 amino acid sequences, identified in the
CC specification (available on a computer readable format), or its fragment,
CC expressed from 2469 of 2469 identified DNA coding regions from the
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
CC AB556454. Also included are an antibody which binds one of the
CC proteins, treating a patient by administering the protein, DNA or
CC antibody (in a composition), a kit comprising first and second primers,
CC which are the nucleic acid cited above or fragments between nucleotides
CC 8-100 of a sequence not defined in the specification, for amplifying a
CC target sequence contained within a Streptococcus nucleic acid sequence,
CC where the first primer is substantially complementary to the target
CC sequence and the second primer is substantially complementary to the
CC complement of the target sequence, and where the parts of the primers
CC having substantial complementarity define the termini of the target
CC sequence to be amplified, assay comprising contacting a test compound
CC with the protein, and determining whether the test compound binds to the
CC protein and a Streptococcus pneumoniae bacterium, where one or more
CC genes encoding the proteins has been rendered inactive. The proteins,
CC nucleic acid molecules, antibody and compositions are useful as
CC medicaments for treating or preventing a disease or infection due to
CC streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,
CC sepsis, otitis media or ear infection. They are also useful in developing
CC vaccines, diagnostics and antibiotics. The methods are useful for
CC identifying immunodominant proteins. The present sequence is one of
CC the 2469 proteins expressed by the identified coding regions from the
CC genomic sequence.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX XX
SQ Sequence 2140 AA;

```

Query Match 99.6%; Score 4380.5; DB 24; Length 2140;

Best Local Similarity 99.6%; Pred. No. 1.9e-62; Matches 664; Conservative 0; Mismatches 0; Indels 3; Gaps 2;

```

QY 1 Y-P--VVLADTSSSEDLNLSDEKEVAENKEKHENISAMETSODFEKTAIVIKEEV 57
DB 19 YAPNEVLAOTSSSEDLNLSDEKEVAENKEKHENISAMETSODFEKTAIVIKEEV 78
QY 58 SKNPVIDNNTSNEBAKIKEENSNKSGDYDTSFVNKNTENPKEDKYVYIAEFKDSGE 117
DB 79 SKNPVIDNNTSNEBAKIKEENSNKSGDYDTSFVNKNTENPKEDKYVYIAEFKDSGE 138
QY 118 KAIKELSSLKTKLYYDRIFNGSAIETTPDNLDKIKOIEGISVERAOKVQPMNHAR 177
DB 139 KAIKELSSLKTKLYYDRIFNGSAIETTPDNLDKIKOIEGISVERAOKVQPMNHAR 198
QY 178 KEIGVEEAIDYLKSIINAFPGNFDGRGVISNIDTGTDIRRHKARIDDDAKASRPFKED 237
DB 199 KEIGVEEAIDYLKSIINAFPGNFDGRGVISNIDTGTDIRRHKARIDDDAKASRPFKED 258
QY 238 LKGTDKXWYLSDKIPHAFFNYNGSKITVEKYDDGRDYFDPHGMIAGIAGLNDTEODIKN 297
DB 259 LKGTDKXWYLSDKIPHAFFNYNGSKITVEKYDDGRDYFDPHGMIAGIAGLNDTEODIKN 318
QY 298 FNGIDGIAPNAQISYKMYSDAGSGFAGDETFMFAIEDSIKHNVDVSVSSGFTGTGLVG 357
DB 319 FNGIDGIAPNAQISYKMYSDAGSGFAGDETFMFAIEDSIKHNVDVSVSSGFTGTGLVG 378
QY 358 EKYWQAIRALRKAGIPMVVATGNVATSASSSSMDLVANNHLMKMTDGTGNVTRTAHEDAIA 417
DB 379 EKYWQAIRALRKAGIPMVVATGNVATSASSSSMDLVANNHLMKMTDGTGNVTRTAHEDAIA 438
QY 418 VASAKNOTVEFDKVNIGESFKYRNIGAFPDKSKITTNEDGTAPSKLKFYIGKGODD 477
DB 439 VASAKNOTVEFDKVNIGESFKYRNIGAFPDKSKITTNEDGTAPSKLKFYIGKGODD 498

```

QY 478 LIGDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVNTVNYNRDNWTELPAWGYZA 537  
 DB 499 LIGDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVNTVNYNRDNWTELPAWGYZA 558  
 QY 538 DEGTQSQVFSISGDDGVKLNWNNINPDKTEVRRNKEDFKDLEQYYPIDMESFNSNKP 597  
 DB 559 DEGTQSQVFSISGDDGVKLNWNNINPDKTEVRRNKEDFKDLEQYYPIDMESFNSNKP 618  
 QY 598 VGDEKEIDFKAPDPTDKELYKEDIIVPAGSTSWGPRIDLLKPDVSAFGKNIKSTLNVIN 657  
 DB 619 VGDEKEIDFKAPDPTDKELYKEDIIVPAGSTSWGPRIDLLKPDVSAFGKNIKSTLNVIN 678  
 QY 658 GKSTYGY 664  
 DB 679 GKSTYGY 685

RESULT 4  
 AAW61246  
 ID AAW61246 standard; Protein; 1007 AA.

XX  
 AC AAW61246;

DT 02-OCT-1998 (first entry)

XX Streptococcus pneumoniae SPI22 protein.

DE Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;  
 KW detection; pneumonia; otitis media; meningitis.

XX Streptococcus pneumoniae.

XX WO9818930-A2.

XX 07-MAY-1998.

PF 30-OCT-1997; 97WO-US19422.

PR 31-OCT-1996; 96US-0029960.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Choi GH, Hromockyj A, Johnson LS, Kunsch CA;

XX WPI; 1998-272224/24.

DR N-PSDB; AAV27431.

XX Nucleic acid encoding antigenic peptide(s) from Streptococcus  
 PT pneumoniae - or their epitope-containing fragments, useful in  
 PT protective or therapeutic vaccines, and for diagnosis

XX Claim 11; Page 92-93; 118pp; English.

XX The present sequence represents a protein from Streptococcus pneumoniae.  
 CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein  
 CC can be useful in vaccines for inducing protective antibodies against  
 CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.  
 CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid  
 CC are used to detect Streptococcus infection (by usual hybridisation or  
 CC amplification methods), also for isolating Streptococcus genes or their  
 CC allelic variants. The protein can be used similarly to detect specific  
 CC antibodies in standard immunoassays, especially for diagnosing or  
 CC monitoring infections. Antibodies which bind the protein are used to  
 CC detect corresponding antigens, to purify the protein and for passive  
 CC immunisation (optionally coupled to a toxin). Vaccines are administered,  
 CC e.g. by injection, orally or through the skin, typically at 0.01-1000  
 CC (especially 10-300) mu g/ml per dose.

XX Sequence 1007 AA;

Query Match 94.8%; Score 4171; DB 19; Length 1007;  
 Best Local Similarity 100.0%; Pred. No. 8e-60; PA  
 Matches 627; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 ETSQDFEKKTAVTKEKEVSKNPVIDNNTSNEBAKIKEENSKSQGYTDSFVNKNTEN 97  
 DB 1 ETSQDFEKKTAVTKEKEVSKNPVIDNNTSNEBAKIKEENSKSQGYTDSFVNKNTEN 60  
 QY 98 PKCEDKVVYIAEFKDKESGEKAIKELSSLKNTKVLTYDRIFNLSAIEETTPDNLDKIKQI 157  
 DB 61 PKCEDKVVYIAEFKDKESGEKAIKELSSLKNTKVLTYDRIFNLSAIEETTPDNLDKIKQI 120  
 QY 158 EGISSEVERAQVQPMNHARKEIGVEEADYLSINAPFGKNFQGRGMVINSIDTGTDYR 217  
 DB 121 EGISSEVERAQVQPMNHARKEIGVEEADYLSINAPFGKNFQGRGMVINSIDTGTDYR 180  
 QY 218 HKAMRIDDDAKASMRFKKEDLKGTDKNYWLSDKIPHAFNYNGGKITVEKYDDGDRYDPP 277  
 DB 181 HKAMRIDDDAKASMRFKKEDLKGTDKNYWLSDKIPHAFNYNGGKITVEKYDDGDRYDPP 240  
 QY 278 HGMHIAGILAGNDEQDIKNFNGIDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSI 337  
 DB 241 HGMHIAGILAGNDEQDIKNFNGIDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSI 300  
 QY 338 KHNVDVSVSGFTGTGLVGEKYQWQAIRALKRAGIPMVVATGNYATSSSSSWDLVANNH 397  
 DB 301 KHNVDVSVSGFTGTGLVGEKYQWQAIRALKRAGIPMVVATGNYATSSSSSWDLVANNH 360  
 QY 398 LKMTDTGNVTRTAHEDAIASAKNOTVEFDKVNIGGESFKYRNIGAFFDKSKITTTNED 457  
 DB 361 LKMTDTGNVTRTAHEDAIASAKNOTVEFDKVNIGGESFKYRNIGAFFDKSKITTTNED 420  
 QY 458 GTKAPSKLKFVYIGKQDQDLIGDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVYN 517  
 DB 421 GTKAPSKLKFVYIGKQDQDLIGDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVYN 480  
 QY 518 TVNYYNRDNWTELPAMGYEADGKTSQVFSISGDDGVKLNWNNINPDKTEVRRNKEDPK 577  
 DB 481 TVNYYNRDNWTELPAMGYEADGKTSQVFSISGDDGVKLNWNNINPDKTEVRRNKEDPK 540  
 QY 578 DKLEQYYPIDMESFNSNKPNGDEKEIDFKAPDPTDKELYKEDIIVPAGSTSWGPRIDLL 637  
 DB 541 DKLEQYYPIDMESFNSNKPNGDEKEIDFKAPDPTDKELYKEDIIVPAGSTSWGPRIDLL 600  
 QY 638 LKPDVSAFGKNIKSTLNIVINGKSTYGY 664  
 DB 601 LKPDVSAFGKNIKSTLNIVINGKSTYGY 627

RESULT 5

ABP54664

ID ABP54664 standard; Protein; 1007 AA.

XX AC ABP54664;

XX DT 04-SEP-2002 (first entry)

XX DE S. pneumoniae SPI22 protein sequence SEQ ID NO:216.

XX KW Streptococcus pneumoniae; epitope; vaccine; antigenic protein;  
 XX antibacterial; Streptococcal infection; detection.

XX OS Streptococcus pneumoniae.

XX PN US2002061545-A1.

XX PD 23-MAY-2002.

XX PF 22-JAN-2001; 2001US-0765272.

XX PR 30-OCT-1997; 97US-0961083.

XX PA (CHOI/) CHOI G H.

XX PA (KUNS/) KUNSCH C A.

XX PA (BARA/) BARASH S C.

XX PA (DILL/) DILLON P J.



PA (DOUG/) DOUGHERTY B.  
 PA (FANN/) FANNON M R.  
 PA (ROSE/) ROSEN C A.  
 XX  
 PI Choi GH, Kunech CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;  
 PI Rosen CA;  
 XX  
 DR WPI: 2002-479261/51.  
 DR N-PSDB; ABQ84899.  
 XX  
 PT New Streptococcus pneumoniae antigens, useful for detecting  
 PT Streptococcus and for preventing or attenuating disease caused by  
 PT Streptococcus infection -  
 XX  
 PS Claim 11; Page 50; 70pp; English.  
 XX  
 CC ABQ84792 to ABQ84904 represents nucleic acids which encode the  
 CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669.  
 CC The S. pneumoniae antigens have antibacterial activity and can be  
 CC used in vaccines. The S. pneumoniae antigens can also be used to  
 CC prevent or attenuate a Streptococcal infection in an animal. The  
 CC polynucleotides encoding the S. pneumoniae antigens can be used to  
 CC detect Streptococcus nucleic acids. ABQ84905 to ABQ85130 represent  
 CC primers used in the cloning of S. pneumoniae ORFs (open reading frames)  
 CC which are used in an example from the present invention.  
 CC  
 XX Sequence 1007 AA;  
 SO  
 Query Match 94.8%; Score 4171; DB 23; Length 1007;  
 Best Local Similarity 100.0%; Pred. No. 8e-60;  
 Matches 627; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 38 ETSODPFKEKTAIVKEKEVSKNPIDNNTSNEAKTEENSOGGYTDSFVNKNTEN 97  
 DB 1 ETSODPFKEKTAIVKEKEVSKNPIDNNTSNEAKTEENSOGGYTDSFVNKNTEN 60  
 QY 98 PKKEDKVVYIAEFKDKESGEKAIKELSSLKNTKVLYTDRIFNCSAIEETPDNLDKIKOI 157  
 DB 61 PKKEDKVVYIAEFKDKESGEKAIKELSSLKNTKVLYTDRIFNCSAIEETPDNLDKIKOI 120  
 QY 158 EGISSEVERAQQVQPMNHARKEIGVEAIDVYKLSINAPFGKNPDGRGVNISNIDGTDYR 217  
 DB 121 EGISSEVERAQQVQPMNHARKEIGVEAIDVYKLSINAPFGKNPDGRGVNISNIDGTDYR 180  
 QY 218 HKARIDDDAKASRFKEDLKGTDKNWLSDKIPIHANNYNGKITTEKXDDGRDYDP 277  
 DB 181 HKARIDDDAKASRFKEDLKGTDKNWLSDKIPIHANNYNGKITTEKXDDGRDYDP 240  
 QY 278 HGMHAGILAGNDETDIKFNFGIDGIAIPNAQIFSYKNYSDAGSGFAGDETFHAIEDSI 337  
 DB 241 HGMHAGILAGNDETDIKFNFGIDGIAIPNAQIFSYKNYSDAGSGFAGDETFHAIEDSI 300  
 QY 338 KHNVVVSVSSGFTGTGLVGEKXWQAIRALKRAGIPMVVATGVNTVYTSASSSSMDLVANNH 397  
 DB 301 KHNVVVSVSSGFTGTGLVGEKXWQAIRALKRAGIPMVVATGVNTVYTSASSSSMDLVANNH 360  
 QY 398 LKMTDGTGVNTTAHEDIAAASAKNQVVEPDKNIGSESPRYRITGAFPFKSKITTMED 457  
 DB 361 LKMTDGTGVNTTAHEDIAAASAKNQVVEPDKNIGSESPRYRITGAFPFKSKITTMED 420  
 QY 458 GTKAPSKLKFVYIGKQDODLIGDLRGKIAVMRIYTKDLKNAFKKAMDGARAIVYN 517  
 DB 421 GTKAPSKLKFVYIGKQDODLIGDLRGKIAVMRIYTKDLKNAFKKAMDGARAIVYN 480  
 QY 518 TVNYNNDNMTLPMWGYEADGTSQVFSISGDDGVKLMNINDPDKTEVYRNNKEDFK 577  
 DB 481 TVNYNNDNMTLPMWGYEADGTSQVFSISGDDGVKLMNINDPDKTEVYRNNKEDFK 540  
 QY 578 DKLQGYVIDMESFNSKNPNVGEDEKEIDKFAPODKLKYKEDIIVPAGSTWGRIDLL 637  
 DB 541 DKLQGYVIDMESFNSKNPNVGEDEKEIDKFAPODKLKYKEDIIVPAGSTWGRIDLL 600  
 QY 638 LKPDVSAFGKNIKSTLVNINCKSTYGY 664

DB 601 LKPDVSAFGKNIKSTLVNINCKSTYGY 627  
 |||  
 RESULT 6  
 ID AAY86159 standard; Protein; 192 AA.  
 XX  
 AC AAY86159;  
 XX  
 DT 10-APR-2000 (first entry)  
 XX  
 DE S. pneumoniae derived protein #368.  
 XX  
 KW Treatment; prevention; disease; diagnosis; gene therapy; screening;  
 KW bacterial; antimicrobial; antibiotic; pathogenesis; infection.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 PN WO9806734-A1.  
 XX  
 PD 19-FEB-1998.  
 XX  
 PF 15-AUG-1997; 97WO-US14436.  
 XX  
 PR 16-AUG-1996; 96US-0024022.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 XX  
 PI Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;  
 PI Stodola RK;  
 XX  
 DR WPI: 1998-159452/14.  
 DR N-PSDB; AAZ96473.  
 XX  
 PT Streptococcus pneumoniae proteins and related DNA - useful for  
 PT screening compounds for antibacterial activity  
 XX  
 PS Claim 5; Page 617; 640pp; English.  
 XX  
 CC This invention describes novel isolated Streptococcus pneumoniae  
 CC polynucleotides (see AAZ96173-296494) and their encoded proteins (see  
 CC AAZ85792-Y86182). The DNA, vectors and host cells described in the  
 CC method of the invention are useful for the recombinant expression of the  
 CC polypeptides. The polypeptides are useful for treatment or prevention of  
 CC disease, or diagnosis of disease related to expression or activity of  
 CC such a polypeptide. They can also be used to screen for compounds which  
 CC interact with and inhibit or activate such a polypeptide. The  
 CC polypeptides (or DNA encoding them, via gene therapy) are also useful  
 CC for inducing an immunological response in a mammal. The antagonists are  
 CC useful to inhibit such bacterial polypeptides. The polypeptides are  
 CC particularly useful to identify antimicrobial compounds and antibiotics.  
 CC They are also useful to determine their role in pathogenesis of  
 CC infection, dysfunction and disease.  
 CC  
 XX Sequence 192 AA;  
 SO  
 Query Match 22.2%; Score 975.5; DB 19; Length 192;  
 Best Local Similarity 93.7%; Pred. No. 3.3e-08;  
 Matches 163; Conservative 1; Mismatches 7; Indels 3; Gaps 2;  
 QY 1 Y-P--VVLADSSSEDALNLSDEKVAENKEKHINHSAMETSDPFKEKTAIVKEKEV 57  
 DB 19 YAPNEVVLADSSSEDALNLSDEKVAENKEKHINHSAMETSDPFKEKTAIVKEKEV 78  
 QY 58 SKNPVINDNTSNEBAKIKGSGNSOGDYTDSFVNKNTENPKKEDKVVYIAEFKDKESGE 117  
 DB 79 SKNPVINDNTSNEBAKIKGSGNSOGDYTDSFVNKNTENPKKEDKVVYIAEFKDKESGE 138  
 QY 118 KAIKELSSLKNTKVLYTDRIFNCSAIEETPDNLDKIKQIEGISSVERAQQVOP 171  
 DB 139 KAIKELSSLKNTKVLYTDRIFNCSAIEETPDNLDKIKQIEGISSVERAQQVHP 192

```

RESULT 7
ID AAR10559 standard; Protein; 1962 AA.
XX
XX AAR10559;
XX
DT 15-APR-1991 (first entry)
XX
DE Mutant protease (A137G/K138L/T139A).
XX
DE Mutant protease gene; fermentation; foodstuff; flavouring;
KW lactic acid bacteria.
XX
OS Lactococcus lactis SK11.
XX
XX Key Location/Qualifiers
FH Peptide 1..187
FT Protein /label= signal_peptide
FT Protein 188..1962 /label= mature_protein
XX
XX EP411715-A.
XX
XX 06-FEB-1991.
XX
XX 02-AUG-1990; 90EP-0202113.
XX
XX 04-AUG-1989; 89NL-0002010.
XX
XX (NEZU-) NED INST ZUIVELONDE.
XX
XX Vos PAJ, Siezen RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;
XX WPI; 1991-038622/06.
XX N-PSDB; AAQ10413.
XX
XX Mutant protease gene(s) and protease(s) - derived from type I and
XX III protease genes from lactococcal strains, used in fermentation
XX foodstuffs and flavourings
XX
XX Disclosure; Fig 1(1-7)+5(b); 29 pp; English.
XX
XX The mutant protease having new cleavage specificities is
XX obtained by replacing three amino acids.
XX This mutant may then be used to prepare hybrid proteases,
XX the fusion being between a type I and a type III protease of
XX L.lactis Wg2 and SK11.
XX The product has modified properties, e.g. thermostability,
XX alkaline/acid pH stability, oxidative stability, autolysis etc.,
XX compared to the parent protease(s). The proteases can be used for
XX preparing products (butter cheese, human and animal foodstuffs)
XX prepared with the aid of lactic acid bacteria.
XX See also AAQ10411-17 and AAQ10870-71.
XX
XX Sequence 1962 AA;
XX
XX Query Match 21.9%; Score 962.5; DB 12; Length 1962;
XX Best Local Similarity 26.1%; Pred. No. 3.2e-05;
XX Matches 182; Conservative 134; Mismatches 241; Indels 139; Gaps 33;
XX
XX 1 YPV--VLADTSSEDAALNISDEKVAENKEKHENHISAMETSQDPFKEKTAIVKEVVS 58
XX 24 LPVGEIOAKAASQOTLSSANTVTAA-----TAKQAATD-----TTA-----AT 64
XX
XX 59 KNPVINDNNTSNEAKIKENSNGDGYTDSFVNKNTENPKKEDKVYVIAEPKDESGEK 118
XX 65 TNOATATQLAAKIDYNKLVKQVQDDIYVDVIVQM--SAAPASENGILRTDYSSTAEIQE 123
XX
XX 119 AIKELSSLKNTKLVY-----TYDRIFNGSAIETTPDNLDKIKQIEGISSVERAOK 168
XX 124 TNKVIAQAASVKAAEVQVTTQQTAGESYGVVNGFSTKVRVWDIPKLQIAGVKTVTLAKV 183
XX
169 VQPMNHARKEIGVEEA-IDYLKINSAPFGKNFDGRGMVINSINIDTGTDIRHKAMRIDDA 227
184 YPTDAKANSMANVQAVMSNY-----KYKGEIVVSVISIDGIDPTHKDMRLSDDK 233
228 KASMRFKKEDL-KGTDK---NYWLSDKTIPAFNY-YNGGKITVEKYDGRDYDFPHGMHI 282
234 DV--KLTKSDVEKFTDTVKHGRYFNSKPYGFVADNNDTIT----DDKVD--EQHGMHV 285
283 AGILAGNTEODIKNFNGIDGIAPNAQIFSKYKMSYDAG-SGFAGDETWFHATEDSIKHN 341
286 AGIIGANGTGDPA--KSVVGVAPPAQLAMKVFNSDTSGLAGSATVVSIEDSAKIGA 343
342 DVVSVSSG-FTGTGLVGEKYQOAIRALKRAGIPMVVATGNVATSSSSSWDLVANNHLM 400
344 DVLNMSLSGNSGNTLEDPELAAVQVNAESGTAAVISAGNSGTSATGEG--VNKDYIYL 401
401 TDTGNVTRTAHEDAIIVASAKNOTVEFDKVI-----GGESFKYRN---IGAFDDK 449
402 QDNEMVSGSPGTSRGATTVASAENTDVIQAVTITDGTGLQLGPETIQLSSHDFTGS-FDQ 460
450 SKITTNEGTAKPSKLFVYIGKGQDQLIGLDLKGIAVMDRI-YTKDLKNFAFKAMDK 508
461 KKFYIVKDAAGN-----LSKALADYTA-DAKGKIAIVKRGESFDDKQKQAAA-- 509
509 GARAIMVNTVNYNNDNWTLPAMGYEADGCTKSQVFSISGDDGVKL--WNMINPKDKT 566
510 GAAGLIIVNT-----DGTATPMT--SIALTTTPTFGLSSVTQKLVDTWTAHPDDSL 560
567 EVKRNKEDFKDKLE-QYYPIDMESFNKNPNVGDEKIDFKFAPDTPDKELYKEDIIVPA 625
561 GVKIT-----LAMLPNQYTEDKMS-----DF----- 582
626 GSTSHGPRIDLLKDPDSAPGKNIKSTLNVINGKST 661
583 --TSYGPVSNLSFKPDITAPGGINWSTQN--NNGYTN 615

RESULT 8
AAR10941
ID AAR10941 standard; Protein; 1968 AA.
XX
XX AAR10941;
XX
XX 15-APR-1991 (first entry)
XX
XX Mutant protease (delta137-139/ins9).
XX
XX Mutant protease gene; fermentation; foodstuff; flavouring;
KW lactic acid bacteria.
XX
XX Lactococcus lactis SK11.
XX
XX Key Location/Qualifiers
FH Peptide 1..187 /label= sig_peptide
FT Protein 188..1968 /label= mat_protein
XX
XX EP411715-A.
XX
XX 06-FEB-1991.
XX
XX 02-AUG-1990; 90EP-0202113.
XX
XX 04-AUG-1989; 89NL-0002010.
XX
XX (NEZU-) NED INST ZUIVELONDE.
XX
XX Vos PAJ, Siezen RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;
XX WPI; 1991-038622/06.
XX N-PSDB; AAQ10871.
XX

```

PT Mutant protease gene(s) and protease(s) - derived from type I and  
PT III protease genes from lactococcal strains, used in fermentation  
PT foodstuffs and flavourings

XX Disclosure: Fig 1(1-7)+5(b); 29pp; English.

CC The wild-type *L. lactis* SK11 protease gene sequence was determined  
CC by the applicant (EP-307011).

CC The mutant protease having new cleavage specificities is  
CC obtained by deleting three amino acids (nine bps) and inserting  
CC 9 other residues.

CC This mutant may then be used to prepare hybrid proteases,  
CC the fusion being between a type I and a type III protease of  
CC *L. lactis* Wg2 and SK11.

CC The product has modified properties, e.g. thermostability,  
CC alkaline/acid pH stability, oxidative stability, autolysis etc.,  
CC compared to the parent protease(s). The proteases can be used for  
CC preparing products (butter cheese, human and animal foodstuffs)

CC prepared with the aid of lactic acid bacteria.

CC See also AAQ10411-17 and AAQ10870-71.

XX Sequence 1968 AA;

Query Match 21.7%; Score 954.5; DB 12; Length 1968;  
Best Local Similarity 26.1%; Pred. No. 4.4e-05;  
Matches 183; Conservative 133; Mismatches 241; Indels 145; Gaps 35;

```

OY 1 YPV--VLADTSSSDALNISDEKVAENKEHENIHSAMETSODPKKTAIVIKEVVS 58
DB 24 LPVGEIOAKKAISOOTLGSSLANVTAA-----TAKQATD-----TTA-----AT 64
OY 59 KNPVLDNNTSBEAKIKENSNKSGDYTDSPFNKNTENPKKEDKVYVIAEFKDESGEK 118
DB 65 TNQAIATQLAAGIDYNNKLVKQODIYVDIVQM--SAPASENGILTTDVSSTAEIOOE 123
OY 119 AIKELSSLKNTKLVY-----TYDRIFNGSAIETTPNLDKIKOIEGISSVERAOK 168
DB 124 TNKVIAAQAASVAAVEQYTOQTAGESYGVVNGFSTKRVVDIPLKQIAGKVTTLAKV 183
OY 169 VQPMNTHARKIEGVEA-IDYIKSINAPGKNFDRGNVISNIDGTGYRRHKARIDDA 227
DB 184 YVPTDAKANSMAVQAVMSNY-----KYKGEQTVSVISDGIPTHKMRLSDDK 233
OY 228 KASRFKEDL-KGTDK---NYWLSDKI PHAFNY--YNGGKTI VEKDDGROVDFPHGMI 282
DB 234 DV--KLTISDVEKFTDYVHGGRYENSKVPYGFNVADNDTIT---DQKVD--EQHGMV 285
OY 283 AGILAGNDTEODIKNFNGIDGIAFNAQIFSYKMY--SD---AG-SGFAGDETFHAIED 335
DB 286 AGIIGANGTDPPA--KSVVGVAPEAQQLAMKVFPSNDSITSGAGTTGTTGSAITVSAIED 343
OY 336 SIKHNVAVSVSSG-FTGTGLGVEKRYKQAIRLRAGI.PMVVATGVATSVASSSSWDIVA 394
DB 344 SAKIGADVIVNMSIGNSGQGTLEDELAAVQANASGTAIVISAGNSGTSAGTGG--VN 401
OY 395 NNHLKMTDTGVNTRPAHEDAI VASAKQYTFEPKNI-----GGESEKRYN---I 443
DB 402 KDYGLQONEMVSGPSTRGATTVASAEVTDVITQAVTITDGTGLQDLPETITQLSDHPT 461
OY 444 GAFPKSKITTTEDDTKAPSKLFFVYIGKQDODLIGDLRCKIAMVDRI--YTKLKNAF 502
DB 462 GS-FQOKPFYIVKDSGN-----LSKGALADYTA--DAKKIIVRGESFPDCKOXY 511
OY 503 KKAMDKGARAIWVNTVYVNRDNMTL.PANGVEADEGTSQVFSISGDDGVKL--WNMI 560
DB 512 AQA--GAAGLIIYNT-----DGTATPMT--SIALTTFPTFGLSVGTGKLVDWYTA 560
OY 561 NPDKTEYKRNNKEPFKQLE--QYYPIMESFNSKKPNVGDEKEIDFKFAPTDKELYKE 619
DB 561 HPDDSLGVKIT-----LAWLPNQKTTEDKMS-----DF-----588
OY 620 DIVPAGSTSMGPRIDLKLPDVSAFGKNIKSTLVINGKST 661

```

DB 589 -----TSYGPVSNLSFKPDITAPGNIWSTON-NGGYTN 621

RESULT 9  
AARI0557  
ID AARI0557 standard; Protein; 1962 AA.

AC AARI0557;

DT 15-APR-1991 (first entry)

XX Mutant protease (A137G/K138D).

XX Mutant protease gene: fermentation; foodstuff; flavouring;

XX lactic acid bacteria.

OS Lactococcus lactis SK11.

XX Key Location/Qualifiers

FT Peptide 1..187

FT Protein /label= signal\_peptide

FT Protein /label= mature\_protein

PD EP411715-A.

XX 06-FEB-1991.

PF 02-AUG-1990; 90EP-0202113.

PR 04-AUG-1989; 89NL-0002010.

XX (NEZU-) NED INST ZUIVELONDE.

PI Vos PAJ, Siezen RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;

DR WPI; 1991-038622/06.

XX N-PSDB; AAQ10411.

PT Mutant protease gene(s) and protease(s) - derived from type I and

PT III protease genes from lactococcal strains, used in fermentation

PT foodstuffs and flavourings

XX Disclosure: Fig 1(1-7)+5(b); 29 pp; English.

PS The mutant protease A137G/K138D having new cleavage specificities is

XX obtained by replacing two amino acids.

CC This mutant may then be used to prepare hybrid proteases,

CC the fusion being between a type I and a type III protease of

CC *L. lactis* Wg2 and SK11.

CC The product has modified properties, e.g. thermostability,

CC alkaline/acid pH stability, oxidative stability, autolysis etc.,

CC compared to the parent protease(s). The proteases can be used for

CC preparing products (butter cheese, human and animal foodstuffs)

CC prepared with the aid of lactic acid bacteria.

CC See also AAQ10411-17 and AAQ10870-71.

XX Sequence 1962 AA;

Query Match 21.7%; Score 952.5; DB 12; Length 1962;  
Best Local Similarity 26.0%; Pred. No. 4.8e-05;  
Matches 181; Conservative 134; Mismatches 242; Indels 139; Gaps 33;

```

OY 1 YPV--VLADTSSSDALNISDEKVAENKEHENIHSAMETSODPKKTAIVIKEVVS 58
DB 24 LPVGEIOAKKAISOOTLGSSLANVTAA-----TAKQATD-----TTA-----AT 64
OY 59 KNPVLDNNTSBEAKIKENSNKSGDYTDSPFNKNTENPKKEDKVYVIAEFKDESGEK 118
DB 65 TNQAIATQLAAGIDYNNKLVKQODIYVDIVQM--SAPASENGILTTDVSSTAEIOOE 123
OY 119 AIKELSSLKNTKLVY-----TYDRIFNGSAIETTPNLDKIKOIEGISSVERAOK 168

```

```

Db 124 TNKVIAQAASVKAAVEQVQTQTAGESYGVVNGFSTKVRVVDIPKLGKQIAGVKTVTYLA 183
Qy 169 VQPMNHARKEIGVEEA-IDYLSINAPFGKNGFGRGMVSNIDTGTDIRHAKMRIDDDA 227
Db 184 YPTDANKANSANVQAVMSNY-----KYKGEFTVWSVIDSGIDPTHKMRSLSDK 233
Qy 228 KASMRFKKEDL-KGTDK---NYWLSDKTIPHAFNY-YNGGKITVEKYDDGRDYFDPHGMHI 282
Db 234 DV--KLTKSDVEKFTDTVKGGRYFNSKVPYGFYADNNDTIT-----DDKVD--EQHGMHV 285
Qy 283 AGILAGNDTEODIKNFNGIDGIAPNAQIFSYKMYSDAG-SGFAGDETMFHAIEDSIKENV 341
Db 286 AGIIGANGTGDDPA--KSVVGVAPPAQLAMKVFNSDTSGLTGSATVVSIAEDSAKIGA 343
Qy 342 DVSVSSG-FTGTGLVGEKYQWQAIRALKAGIPMVVATGNVATASASSSSDLVANNHLKM 400
Db 344 DVLNMSLGSNSGNTLEDPELAQVQANESGTAAVISAGNSGTSGSATEG--VNRKDYGL 401
Qy 401 TDGTVTRTAHAHDAIAVASAKNOTVEFDKNI-----GGESPKYRN---IGAFEDK 449
Db 402 QDNEMVSGPSTSGATTVASAENTDVIQAVTITDGTGLGLGPETIQLSHDFTGS-PDQ 460
Qy 450 SKITTNEDGTPKAPSKLVYTGKQDODLIGLGRKIAVMDRI-YTKDLKNAFKKAMD 508
Db 461 KKFVIVKDSGN-----LSKGLADYTA-DAKGKIAIVKRGFEFDDKQYQAQA-- 509
Qy 509 GARAIMVNTVNYNRDWNWTELPAMGYEADSGTKSQVFSISGDDGVKL--WNMINPKKT 566
Db 510 GAAGLIIVNT-----DGTATPMT--SIALTTTPTFGLSVTGQKLVDMVTAHPDDSL 560
Qy 567 EVKRNKEDFKDLE-QYYPIDMESFNSKNPVGDEKIDFKFAPDPTDKELYKEDIIVPA 625
Db 561 GVKIT-----LAMPLPNQKYTEDKMS-----DF----- 582
Qy 626 GSTSGWPRIDLLKPDVSPAGKNITKSTLNVIKNGST 661
Db 583 --TSYGPVSNLSFPRDITAPGNIWSTQN--NNGYTN 615

RESULT 10
AAR10940
ID AAR10940 standard; Protein; 1974 AA.
XX
AC AAR10940;
XX
DT 15-APR-1991 (first entry)
XX
DE Mutant protease (delta137-139/ins15).
XX
KW Mutant protease gene; fermentation; foodstuff; flavouring;
KW lactic acid bacteria.
XX
OS Lactococcus lactis SK11.
XX
FH Key Location/Qualifiers
FT Peptide 1..187
FT /label= SIG PEPTIDE
FT Protein 188..1974
FT /label= MAT_PROTEIN
XX
XX
EP 011715-A.
XX
XX 06-FEB-1991.
XX
XX 02-AUG-1990; 90EP-0202113.
XX
XX 04-AUG-1989; 89NL-0002010.
XX
XX (NEZU-) NED INST ZUIVELONDE.
XX
XX Vos PAJ, Siezen RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;
XX WPI; 1991-038622/06.
XX

```

```

DR N-PSDB; AAQ10870.
XX
XX Mutant protease gene(s) and protease(s) - derived from type I and
XX III protease genes from lactococcal strains, used in fermentation
XX foodstuffs and flavourings
XX
XX Disclosure; Fig 1(1-7)+5(b); 29pp ; English.
XX
XX The wild-type L.lactis SK11 protease gene sequence was determined
XX by the applicant (EP-307011).
XX The mutant protease having new cleavage specificities is
XX obtained by deleting three amino acids (nine bps) and inserting
XX 15 other residues.
XX This mutant may then be used to prepare hybrid proteases, of
XX the fusion being between a type I and a type III protease of
XX L.lactis Wg2 and SK11.
XX The product has modified properties, e.g. thermostability,
XX alkaline/acid pH stability, oxidative stability, autolysis etc.,
XX compared to the parent protease(s). The proteases can be used for
XX preparing products (butter cheese, human and animal foodstuffs)
XX prepared with the aid of lactic acid bacteria.
XX See also AAQ10411-17 and AAQ10870-71.
XX
XX Sequence 1974 AA;
XX
Query Match 21.7%; Score 952.5; DB 12; Length 1974;
Best Local Similarity 25.8%; Pred.No.4.8e-05;
Matches 183; Conservative 133; Mismatches 241; Indels 151; Gaps 35;
Qy 1 YPV--VLADTSSSDALNISDKKVAENKEHENIHSAMETSQDPKPKKTAIVKEVVS 58
Db 24 LPVGEIQAKAISQOTLGSSLANVTAA-----TAKQAATD-----TTA-----AT 64
Qy 59 KNPVIDNNTSNEEAKIKEENSKSGDYTDTSFVNKNTENPKKEDKVVIAEPKDESGEK 118
Db 65 TNQAIATQLAAKGIDYNKLNKVVQODIYVDVIVQM-SAAPASENGILRTDYSSTAIEIQE 123
Qy 119 AIKELSSLKNTKLY-----TYDRIFNGSAIETTPDNLKIKQIEGISSVERAQK 168
Db 124 TNKVIAQAASVKAAVEQVQTQTAGESYGVVNGFSTKVRVVDIPKLGKQIAGVKTVTYLA 183
Qy 169 VQPMNHARKEIGVEEA-IDYLSINAPFGKNGFGRGMVSNIDTGTDIRHAKMRIDDDA 227
Db 184 YPTDANKANSANVQAVMSNY-----KYKGEFTVWSVIDSGIDPTHKMRSLSDK 233
Qy 228 KASMRFKKEDL-KGTDK---NYWLSDKTIPHAFNY-YNGGKITVEKYDDGRDYFDPHGMHI 282
Db 234 DV--KLTKSDVEKFTDTVKGGRYFNSKVPYGFYADNNDTIT-----DDKVD--EQHGMHV 285
Qy 283 AGILAGNDTEODIKNFNGIDGIAPNAQIFSYKMY--SD-----AG-----SGFAGDETM 329
Db 286 AGIIGANGTGDDPA--KSVVGVAPPAQLAMKVFNSDTSGLTGSATVVSIAEDSAKIGA 343
Qy 342 DVSVSSG-FTGTGLVGEKYQWQAIRALKAGIPMVVATGNVATASASSSSDLVANNHLKM 400
Db 344 DVLNMSLGSNSGNTLEDPELAQVQANESGTAAVISAGNSGTSGSATEG--VNRKDYGL 401
Qy 401 TDGTVTRTAHAHDAIAVASAKNOTVEFDKNI-----GGESPKYRN---IGAFEDK 449
Db 402 QDNEMVSGPSTSGATTVASAENTDVIQAVTITDGTGLGLGPETIQLSHDFTGS-PDQ 460
Qy 450 SKITTNEDGTPKAPSKLVYTGKQDODLIGLGRKIAVMDRI-YTKDLKNAFKKAMD 508
Db 461 KKFVIVKDSGN-----LSKGLADYTA-DAKGKIAIVKRGFEFDDKQYQAQA-- 509
Qy 509 GARAIMVNTVNYNRDWNWTELPAMGYEADSGTKSQVFSISGDDGVKL--WNMINPKKT 566
Db 510 GAAGLIIVNT-----DGTATPMT--SIALTTTPTFGLSVTGQKLVDMVTAHPDDSL 560
Qy 567 EVKRNKEDFKDLE-QYYPIDMESFNSKNPVGDEKIDFKFAPDPTDKELYKEDIIVPA 625
Db 561 GVKIT-----LAMPLPNQKYTEDKMS-----DF----- 582
Qy 626 GSTSGWPRIDLLKPDVSPAGKNITKSTLNVIKNGST 661
Db 583 --TSYGPVSNLSFPRDITAPGNIWSTQN--NNGYTN 615

AAR10940
ID AAR10940 standard; Protein; 1974 AA.
XX
AC AAR10940;
XX
DT 15-APR-1991 (first entry)
XX
DE Mutant protease (delta137-139/ins15).
XX
KW Mutant protease gene; fermentation; foodstuff; flavouring;
KW lactic acid bacteria.
XX
OS Lactococcus lactis SK11.
XX
FH Key Location/Qualifiers
FT Peptide 1..187
FT /label= SIG PEPTIDE
FT Protein 188..1974
FT /label= MAT_PROTEIN
XX
XX
EP 011715-A.
XX
XX 06-FEB-1991.
XX
XX 02-AUG-1990; 90EP-0202113.
XX
XX 04-AUG-1989; 89NL-0002010.
XX
XX (NEZU-) NED INST ZUIVELONDE.
XX
XX Vos PAJ, Siezen RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;
XX WPI; 1991-038622/06.
XX

```

Qy 614 KELYKEDIIIVAGSTSGPRIDLLKPDVSAPGKNIKSTLNVINGKST 661  
 Db 595 -----TSYGPVSNLSFKPDITAPGGINIMSTON--NNGYTN 627

## RESULT 11

AA010558  
 ID AAR10558 standard; Protein; 1962 AA.

AC AAR10558;

DT 15-APR-1991 (first entry)

DE Mutant protease (A137G/K138P/T139P).

KM Mutant protease gene; fermentation; foodstuff; flavouring;

KW lactic acid bacteria.

OS Lactococcus lactis SK11.

XX Key Location/Qualifiers

FT Peptide 1..187

FT Protein /label= signal\_peptide

FT Protein /label= mature\_protein

PN EP411715-A.

PD 06-FEB-1991.

PF 02-AUG-1990; 90EP-0202113.

PR 04-AUG-1989; 89NL-0002010.

XX (NEZU-) NED INST ZUIVELONDE.

PI Vos PAJ, Siezen RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;

XX MPI; 1991-038622/06.

DR N-PSDB; AAQ10412.

XX Mutant protease gene(s) and protease(s) - derived from type I and  
 PT III protease genes from lactococcal strains, used in fermentation  
 PT foodstuffs and flavourings

XX Die closure; Fig 1(1-7)+5(b); 29 pp; English.

XX The mutant protease having new cleavage specificities is

CC obtained by replacing three amino acids.

CC This mutant may then be used to prepare hybrid proteases,

CC the fusion being between a type I and a type III protease of

CC L.lactis Wg2 and SK11.

CC The product has modified properties, e.g. thermostability,

CC alkaline/acid pH stability, oxidative stability, autoprocolysis etc.,

CC compared to the parent protease(s). The proteases can be used for

CC preparing products (butter cheese, human and animal foodstuffs)

CC prepared with the aid of lactic acid bacteria.

CC See also AAQ10411-17 and AAQ10870-71.

CC Sequence 1962 AA;

Qy Query Match 21.6%; Score 951.5; DB 12; Length 1962;

Db Best Local Similarity 26.0%; Pred. No. Se-05; Indels 139; Gaps 33;

Matches 181; Conservative 133; Mismatches 243;

1 YPV--VLADTSSSEDALNISDKKAVENKKEHINHSAMETSQDFKKTAVIKKEKVS 58

24 LPVGEIQKAKAISQOFTLSSLANVTYAA-----TAKQAATD-----TTA-----AT 64

59 KNPVIDNNTSNEAKIKKEENSNNKSGQDYTDSFVNKNTNPKKEDKVVYIAEFKKESEK 118

65 TNOAIAITOLAAKIGIDYNNKLNKVKQODIYDVIVQW--SAAPASENGILRTDYSSTAIEIOE 123

Qy 119 AIKELSLKNTKVLV-----TYDRIFNAGSAIETTPDNLKIQIEGIESVERAOK 168  
 Db 124 TNKVIYAQASVKAIVEGVTQGTAGESYGVVNGFSTKVRVVDIPKIQIAGVKTIVLAKV 183

Qy 169 VQPMNHARKKEIGVEEA-IDYLSKINAPFGNPFGRGVINISITGDTYRKKARIDDA 227

Db 184 YPTDAKANSMANVOAAMSNY-----KYNGEGTVASVDSGIDPTKDKRLSDDK 233

Qy 228 KASRPFKEEDL-KGTDK--NYWLSDKIPHAFNY-YNGKKTVEKYDDGRDYPFHGMHI 282

Db 234 DV--KLTKSDVEKFTDVKHGRYNSKVPYGFNTADNNDIT--DDKVD--EONGHNV 285

Qy 283 AGILAGNDTEODIXNFNGIDGIAPNQOIFSYKMYSDAG-SGFAGDETFHAIEDSIXANV 341

Db 286 AGIIGANGTGDDPA--KSVGVAPAEQDLAMKVPNSDTSGPSATATVSAIEDSAKIGA 343

Qy 342 DVSVSSG-FTGTGLVGEKTVQAIRLRKAGIPNVVATGNYATSSSSWDLVANNHLKM 400

Db 344 DVLNMSISGNSGNOTLEDPELAAVQANNESTAAVISAGNSGTSATGEG--VKKDYGL 401

Qy 401 TDTGNVTRTAHEDAIVASAKNOTVERDKNI-----GGSFKYRN--IGAFEDK 449

Db 402 QDNEMVSGPSTSRGATTVASAEHTDVITQAVTITDGTGLQIGPFTIQLSHDFTGS-FDQ 460

Qy 450 SKITTNEBGTAPAPSKLKFVYIGKQODDLGLDRGKIAMVDRI-YTKDLNNAFKKAMDK 508

Db 461 KKFYIVADASGN-----LSKGLADYTA-DAKGKAIYKRGSEFSDDKQKXAQAA-- 509

Qy 509 GARAIMVNVTVVYVNRDNTLPLPMGYEADGTYSVFSGISGDGVKL--NMNINPDKKT 566

Db 510 GNAGLIIVNT-----DGTATPMT--SIALTTFPFGLSSVGVOKLVDVTAHPDSSL 560

Qy 567 EVKRNKNEDEPKDKE-QYPIIDMESFNSNKNVNDGKEDPKFAPDTDKELYKEDIIVPA 625

Db 561 GVKIT--LAMPNQYTEDKMS-----DF----- 582

Qy 626 GSTSGPRIDLLKPDVSAPGKNIKSTLNVINGKST 661

Db 583 --TSYGPVSNLSFKPDITAPGGINIMSTON--NNGYTN 615

RESULT 12  
 AAR10560  
 ID AAR10560 standard; Protein; 1962 AA.

AC AAR10560;

DT 15-APR-1991 (first entry)

DE Mutant protease (K138N).

KM Mutant protease gene; fermentation; foodstuff; flavouring;

KW lactic acid bacteria.

OS Lactococcus lactis SK11.

XX Key Location/Qualifiers

FT Peptide 1..187

FT Protein /label= signal\_peptide

FT Protein /label= mature\_protein

PN EP411715-A.

PD 06-FEB-1991.

PF 02-AUG-1990; 90EP-0202113.

PR 04-AUG-1989; 89NL-0002010.

XX (NEZU-) NED INST ZUIVELONDE.

PI Vos PAJ, Siezen RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;

```
XX WPI; 1991-038622/06.
DR N-PSDB; AAQ10414.
XX
XX Mutant protease gene(s) and protease(s) - derived from type I and
PT III protease genes from lactococcal strains, used in fermentation
PT foodstuffs and flavourings
XX
XX Disclosure; Fig 1(1-7)+5(a); 29 pp; English.
XX
XX The mutant protease K138N having new cleavage specificities is
CC obtained by carrying out single amino acid substitutions.
CC This mutant may then be used to prepare hybrid proteases.
CC the fusion being between a type I and a type III protease of
CC L.lactis Wg2 and SK11.
CC The product has modified properties, e.g. thermostability,
CC alkaline/acid pH stability, oxidative stability, autolysis etc.,
CC compared to the parent protease(s). The proteases can be used for
CC preparing products (butter cheese, human and animal foodstuffs)
CC prepared with the aid of lactic acid bacteria.
CC See also AAQ10411-17 and AAQ10870-71.
XX
XX Sequence 1962 AA;
SQ
Query Match 21.6%; Score 948.5; DB 12; Length 1962;
Best Local Similarity 25.9%; Pred. No. 5.6e-05;
Matches 180; Conservative 134; Mismatches 243; Indels 139; Gaps 33;
QY 1 YPV--VLADTSSSEDALNISDKKVAENKHEHNSAMETSQDPKKEKTKAVIKEVVS 58
DB 24 LPVGEIQAKAISQDTGLSSSLANTVTAA-----TAKQAATD-----TTA-----AT 64
QY 59 KNPVIDNNTSNEEAKIKEENSKSGDYTFVFNKNTENPKKEDKVVVIAEFKDESGEK 118
DB 65 TNQAIATQLAAKGIDYNKLNKVVQODIYVDVIVQM--SAAPASENGILRTDYSSTAEIQOE 123
QY 119 AIKELSSLKNTKVLV-----TYDRIFNGSAIETTPDNLKIKQIEGSSVERAOK 168
DB 124 TNKVIAAQAASVKAAEVQVTOQTAGESYGVVNGFSTKVRVVDIPKQIAGVKTVTLAKV 183
QY 169 VQPMNHARKEIGVEEA-IDYLSKNAPFGKNFDCRGMVISNIDTGTDRHKAIRIDDDA 227
DB 184 YPTDANKANSANVQAQVSNY-----KYKGEIVVSVIDSGIDPTHKMRLSDDK 233
QY 228 KASMRFKKEDL-KGPDK--NYWLSDKTIPHAFNY-YNGKKTIVKYDGRDYDPDGHMHI 282
DB 234 DV--KLTKSDEVEKFTDTVKHGRYFNKVPYGFNVADNDTIT-----DDKVD--EQHGMHV 285
QY 283 AGILAGNTEQDIKNFNGIDGIAPNAQIFSKMYSDAG-SGFAGDETFHAIEDSGIKHNV 341
DB 286 AGITGANGTGDDPA--KSVVGVAPQAQLLAKMVFNSDTSANTGATVVSALIEDSAKIGA 343
QY 342 DVSVSSG-FTGTGLVGEKYWQAIRALKRAGIPMVVATGCVNATSASSSSDVLVANNHLM 400
DB 344 DVLNMSLSGNSGNTLEPDLAAVQNAESGTAAVISAGNSTGSAFEG--VNKDYVGL 401
QY 401 TDTGNVTRTAHEDAIIVASAKNQTVEFDKVI-----GGESFKYRN-----IGAFDPK 449
DB 402 QDNVMVSGPSTSRGATTVASAEVTOAVTITDGTGLQLGPETIQLSSHDFTGS-FDQ 460
QY 450 SKITTNEGTRKAPSKLVFYICKGQODLIGLDRGKTAVMDRI-YTKDLKNAFKKAMDK 508
DB 461 KKFYIVKDSAGN-----LSKALADYTA-DAKGKTAIVKRGFSFDDKQKYAQA-- 509
QY 509 GARAIMVNVYNYNRDNWTELPAMGYEADGKTSQVFSISDDGVKL--NMNINPDKKT 566
DB 510 GAAGLIIVNT-----DGTATPMT--SIALTTFPTFGLSSVTGQKLVDMWTAHPDDSL 560
QY 567 EYVRNNKEDFKKLE-QYYPIDMESFNSKNPNVGDEKEIDFKFAPDPTDKELYKEDIIVPA 625
DB 561 GVKIT-----LAMLPNQKYTEDKMS-----DF----- 582
QY 626 GSTNGPRIDLLLLKPDVSAFGNKIKSTLVNKGST 661
```

```
DB 583 --TSYGPVSNLSFKPDITAPGGINWSTQN-NNGYTN 615
XX
XX RESULT 13
XX AAR10561
XX AAR10561 standard; Protein; 1962 AA.
XX
XX AAR10561;
AC
XX 15-APR-1991 (first entry)
DT
XX Mutant protease (N166D).
DE
XX Mutant protease gene; fermentation; foodstuff; flavouring;
KW lactic acid bacteria.
XX
XX Lactococcus lactis SK11.
OS
XX Key Location/Qualifiers
FH 1..187
FT /label= signal_peptide
FT 188..1962
FT Protein
FT /label= mature_protein
XX
XX EP411715-A.
PN
XX 06-FEB-1991.
PD
XX 02-AUG-1990; 90EP-0202113.
PF
XX 04-AUG-1989; 89NL-0002010.
PR
XX (NEZU-) NED INST ZUIVELONDE.
PA
XX Vos PAU, Siezen RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;
PI WPI; 1991-038622/06.
DR N-PSDB; AAQ10415.
XX
XX Mutant protease gene(s) and protease(s) - derived from type I and
PT III protease genes from lactococcal strains, used in fermentation
PT foodstuffs and flavourings
XX
XX Disclosure; Fig 1(1-7)+5(a); 29 pp; English.
XX
XX The mutant protease N166D having new cleavage specificities is
CC obtained by carrying out single amino acid substitutions.
CC This mutant may then be used to prepare hybrid proteases.
CC the fusion being between a type I and a type III protease of
CC L.lactis Wg2 and SK11.
CC The product has modified properties, e.g. thermostability,
CC alkaline/acid pH stability, oxidative stability, autolysis etc.,
CC compared to the parent protease(s). The proteases can be used for
CC preparing products (butter cheese, human and animal foodstuffs)
CC prepared with the aid of lactic acid bacteria.
CC See also AAQ10411-17 and AAQ10870-71.
XX
XX Sequence 1962 AA;
SQ
Query Match 21.6%; Score 948.5; DB 12; Length 1962;
Best Local Similarity 25.9%; Pred. No. 5.6e-05;
Matches 180; Conservative 134; Mismatches 243; Indels 139; Gaps 33;
QY 1 YPV--VLADTSSSEDALNISDKKVAENKHEHNSAMETSQDPKKEKTKAVIKEVVS 58
DB 24 LPVGEIQAKAISQDTGLSSSLANTVTAA-----TAKQAATD-----TTA-----AT 64
QY 59 KNPVIDNNTSNEEAKIKEENSKSGDYTFVFNKNTENPKKEDKVVVIAEFKDESGEK 118
DB 65 TNQAIATQLAAKGIDYNKLNKVVQODIYVDVIVQM--SAAPASENGILRTDYSSTAEIQOE 123
QY 119 AIKELSSLKNTKVLV-----TYDRIFNGSAIETTPDNLKIKQIEGSSVERAOK 168
```

```

Db      124 TNKVIAQASVKAAVEQTQOTAGESYGVVNGFSTKRVVDIPKLQIAGVKTVTLAKV 183
Qy      169 VQPMNNAHKEIGVEA-IDYLSINAPFGNFDGRGVINISITGTDYRHKAMRIDDA 227
Db      184 YPTDAKANSNANVOAVWSNY-----KYRGEQTVVSDIGDPTKMRSLSDK 233
Qy      228 KASRPFKEDL-KGTDK---NYWLSDKIPHAFNY-YNGSKITVEKYDGRDPFPHGMHI 282
Db      234 DV--KLTSDEVEKFTDYVKGGRYNSKVPYGFENADNDIT---DDKVD--EQHGHHV 285
Qy      283 AGILAGNTEODIKNFNGIDGIAPNAQIFSYKMSDAG-SGFAGDETFHAIEDSIKHNV 341
Db      286 AGIIGANGTGDPA--KSVGVAPPAQLAMKMFNSDTSANTGSATVVAIEDSAKIGA 343
Qy      342 DVSVSSGF-TGTGLVEKYMQAIRALKAGIPMVVATGNVATSSSSSWDLVANNHLM 400
Db      344 DVLNMSLGSNGNQTLEDPDLAAVONANESGTAIVASNGSGTSATEG--VKKDYGL 401
Qy      401 TDTGNVTRTAHEDAIIVASAKNQTVEFDKNI-----GGSEFKYRN--IGAFEDK 449
Db      402 QDNEMVSGPTSGRATTVASAENDVITQAVTITDGTGLQGPETIQLSSHDFTGS-FDQ 460
Qy      450 SKITTNEGTAKPSKLKFEVYIGKGGDDDLGLDLRGKIAVNDRI-YTKDLKNAFKKAMD 508
Db      461 KKFYIVKADSGN-----LSKGLADYTA-DAKGIKAIYKRGESFDDKQKVAQA-- 509
Qy      509 GARAIMVNTVNYVNRDNTLPMAGYEADGTSQVPSISGDDGVKL--WNNINPDKKT 566
Db      510 GAAGLIIVNT-----DGTATPMT--SIATLTTFPTGSLSVTGQKLVDTVAHPDLSL 560
Qy      567 EVKRNKKEDFKDLE-QYYPIDMESFNSNKNVGDDEKIDPKFAPDTDKELYKEDIYPA 625
Db      561 GVKIT---LAMLPNQKYTEDKMS-----DF----- 582
Qy      626 GSTWGPRIIDLKPDVASAPGNIKSTLNVINGKST 661
Db      583 --TSYGPSVNSLFPKPDITAPGGINISTQN--NNGYTN 615

RESULT 14
AAR10563
ID AAR10563 standard; Protein; 1962 AA.
XX
AC AAR10563;
XX
DT 15-APR-1991 (first entry)
XX
DE Mutant protease (K748T).
XX
KM Mutant protease gene; fermentation; foodstuff; flavouring;
KW lactic acid bacteria.
XX
OS Lactococcus lactis SK11.
XX
FH Key Location/Qualifiers
FT Peptide 1..187
FT /label= signal_peptide
FT Protein 188..1962
FT /label= mature_protein
XX
PN EPA11715-A.
XX
PD 06-FEB-1991.
XX
PF 02-AUG-1990; 90EP-0202113.
XX
PR 04-AUG-1989; 89NL-0002010.
XX
PA (NEZU-) NED INST ZUIVELONDE.
XX
PI Vos PAJ, Siezen RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;

```

```

DR      WPI; 1991-038622/06.
DR      N-PSDB; AAQ10417.
XX
PT      Mutant protease gene(s) and protease(s) - derived from type I and
PT      III protease genes from lactococcal strains, used in fermentation
PT      foodstuffs and flavourings
XX
PS      Disclosure; Fig 1(1-7)+5(a); 29 pp; English.
XX
CC      The mutant protease K748T having new cleavage specificities is
CC      obtained by carrying out single amino acid substitutions.
CC      This mutant may then be used to prepare hybrid proteases,
CC      the fusion being between a type I and a type III protease of
CC      L.lactis Wg2 and SK11.
CC      The product has modified properties, e.g. thermostability,
CC      alkaline/acid pH stability, oxidative stability, autolysis etc.,
CC      compared to the parent protease(s). The proteases can be used for
CC      preparing products (butter cheese, human and animal foodstuffs)
CC      prepared with the aid of lactic acid bacteria.
CC      See also AAQ10411-17 and AAQ10870-71.
XX
SQ      Sequence 1962 AA;
XX
Query Match 21.6%; Score 948.5; DB 12; Length 1962;
Best Local Similarity 25.9%; Pred. No. 5.6e-05;
Matches 180; Conservative 134; Mismatches 243; Indels 139; Gaps 33;

Qy      1 YPV-VIADTSSSEDALNISDEKVAENKEHNIHSMETSOPFEKKTAVITEKEYVS 58
Db      24 LPVEIGAKKAISQOTLGSSLANVTAA-----TAKQAAVT-----AT 64
Qy      59 KNPYIDNNTSNEAKIKIEENSKSQDYDTSFVNKNTPENPKEDKVVYIAFKOKESGEK 118
Db      65 TNQAIATQIAKKGIDYNNKLVQODIYVDIVM-SAAPASENGILRTDYSSIAETIQE 123
Qy      119 AIKELSLKTKTVLY-----TYDRIFNGSAIEETPDNLKIQIEGIESVERAQK 168
Db      124 TNKVIAQASVKAAVEQTQOTAGESYGVVNGFSTKRVVDIPKLQIAGVKTVTLAKV 183
Qy      169 VQPMNNAHKEIGVEA-IDYLSINAPFGNFDGRGVINISITGTDYRHKAMRIDDA 227
Db      184 YPTDAKANSNANVOAVWSNY-----KYRGEQTVVSDIGDPTKMRSLSDK 233
Qy      228 KASRPFKEDL-KGTDK---NYWLSDKIPHAFNY-YNGSKITVEKYDGRDPFPHGMHI 282
Db      234 DV--KLTSDEVEKFTDYVKGGRYNSKVPYGFENADNDIT---DDKVD--EQHGHHV 285
Qy      283 AGILAGNTEODIKNFNGIDGIAPNAQIFSYKMSDAG-SGFAGDETFHAIEDSIKHNV 341
Db      286 AGIIGANGTGDPA--KSVGVAPPAQLAMKMFNSDTSANTGSATVVAIEDSAKIGA 343
Qy      342 DVSVSSGF-TGTGLVEKYMQAIRALKAGIPMVVATGNVATSSSSSWDLVANNHLM 400
Db      344 DVLNMSLGSNGNQTLEDPDLAAVONANESGTAIVASNGSGTSATEG--VKKDYGL 401
Qy      401 TDTGNVTRTAHEDAIIVASAKNQTVEFDKNI-----GGSEFKYRN--IGAFEDK 449
Db      402 QDNEMVSGPTSGRATTVASAENDVITQAVTITDGTGLQGPETIQLSSHDFTGS-FDQ 460
Qy      450 SKITTNEGTAKPSKLKFEVYIGKGGDDDLGLDLRGKIAVNDRI-YTKDLKNAFKKAMD 508
Db      461 KKFYIVKADSGN-----LSKGLADYTA-DAKGIKAIYKRGESFDDKQKVAQA-- 509
Qy      509 GARAIMVNTVNYVNRDNTLPMAGYEADGTSQVPSISGDDGVKL--WNNINPDKKT 566
Db      510 GAAGLIIVNT-----DGTATPMT--SIATLTTFPTGSLSVTGQKLVDTVAHPDLSL 560
Qy      567 EVKRNKKEDFKDLE-QYYPIDMESFNSNKNVGDDEKIDPKFAPDTDKELYKEDIYPA 625
Db      561 GVKIT---LAMLPNQKYTEDKMS-----DF----- 582
Qy      626 GSTWGPRIIDLKPDVASAPGNIKSTLNVINGKST 661

```



```

Db      583 --TSYGPVSNLSFKPDITAPGGNIWSTQN-NGGYTN 615

RESULT 15
AA010562
ID  AA010562 standard; Protein; 1959 AA.
XX
XX
AC  AA010562;
XX
XX  15-APR-1991 (first entry)
DT
XX
XX  Mutant protease (delta137-139).
DE
XX
XX  Mutant protease gene; fermentation; foodstuff; flavouring;
KW  lactic acid bacteria.
XX
XX  Lactococcus lactis SK11.
OS
XX
XX  Key Location/Qualifiers
FH  Peptide 1..187
FT  /label= signal_peptide
FT  188..1959
FT  /label= mature_peptide
XX
XX  EP411715-A.
XX
XX  06-FEB-1991.
PD
XX
XX  02-AUG-1990; 90EP-0202113.
PF
XX
XX  04-AUG-1989; 89NL-0002010.
PR
XX
XX  (NEZU-) NED INST ZUIVELONDE.
PA
XX
XX  Vos PAJ, Siezen RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;
PI  WPI; 1991-038622/06.
XX  N-PSDB; AAQ10416.
DR
XX
XX  Mutant protease gene(s) and protease(s) - derived from type I and
PT  III protease genes from lactococcal strains, used in fermentation
PT  foodstuffs and flavourings
PS  Disclosure; Fig 1(1-7)+5(b); 29 pp; English.
XX
XX  The mutant protease having new cleavage specificities is
CC  obtained by deleting three amino acids.
CC  This mutant may then be used to prepare hybrid proteases,
CC  the fusion being between a type I and a type III protease of
CC  L.lactis Wg2 and SK11.
CC  The product has modified properties, e.g. thermostability,
CC  alkaline/acid pH stability, oxidative stability, autolysis etc.,
CC  compared to the parent protease(s). The proteases can be used for
CC  preparing products (butter cheese, human and animal foodstuffs)
CC  prepared with the aid of lactic acid bacteria.
CC  See also AAQ10411-17 and AAQ10870-71.
XX
XX  Sequence 1959 AA:
SQ

```

```

Query Match      21.5%; Score 947; DB 12; Length 1959;
Best Local Similarity 25.8%; Pred. No. 5.9e-05;
Matches 179; Conservative 135; Mismatches 241; Indels 140; Gaps 33;

QY  1 YPV--VLADTSSSEDALNISDEKVAENKEKHENIHSAMETSQDFKEKKTAVIKEKWS 58
DB  24 LPVGEIQAKAAISQQTIGSSSLANTVTAA-----TAKQAATD-----TTA-----AT 64
QY  59 KNPVINDNTSNEAKIKEENSNSQGDYDTSFVNKNTENPKKEDKVVVIAEFKDKESKEK 118
DB  65 TNQAIATQAAKGIIDYNKLNKVKQQQDIYVDVIVQM-SAAPASENGILRTDYSSTAEIQOE 123
QY  119 AIKELSSLNKTkVLY-----TYDRIFNGSAIETTPDNLDDKIKQIEGISSVERAQK 168

```

Search completed: October 9, 2003, 15:51:10  
Job time : 52 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: October 9, 2003, 15:48:38 ; Search time 85 Seconds

(without alignments)  
1239.935 Million cell updates/sec

Title: US-09-590-991-6

Perfect score: 4398

Sequence: 1 YPVVADTSSSEDALNISDK.....FGKNIKSTLVINGKSTGY 664

Scoring table:

BLOSUM30  
Gapop 8.0 , Gapext 0.1

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

A\_Geneseq\_19Jun03.\*  
1: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4398	100.0	664	22 AAB48342	S. pneumoniae Sp12
2	4398	100.0	2120	21 AAY81710	Streptococcus pneu
3	4381.7	99.6	2140	24 ABU01020	S. pneumoniae type
4	4171	94.8	1007	19 AAM61246	Streptococcus pneu
5	4171	94.8	1007	23 ABP54664	S. pneumoniae Sp12
6	1188.3	27.0	2366	17 AAM68388	Clostridium diffic
7	1188.3	27.0	2366	19 AAR95011	C. difficile toxin
8	1186.4	27.0	3132	22 AAB51367	Japanese yam mosai
9	1175.3	26.7	6619	22 ABB23329	Novel human diagno

10	1173	26.7	5909	22 ABB23295	Novel human diagno
11	1167.8	26.6	7107	23 ABB58144	Drosophila melanog
12	1165.4	26.5	2835	22 ABB98574	Dextran saccharase
13	1165.2	26.5	31267	24 ABB74786	Human RGS11 protei
14	1164.4	26.5	4345	22 ABB66417	Drosophila melanog
15	1161.5	26.4	26926	22 AAU05396	Human titlin (cone
16	1156.9	26.3	2946	22 ABB25790	Novel human diagno
17	1156.1	26.3	5464	22 ABB11810	Novel human diagno
18	1153.3	26.2	3781	22 ABB373291	Novel human diagno
19	1151.9	26.2	4643	22 ABB71609	Drosophila melanog
20	1151.8	26.2	2710	17 AAR35016	C. difficile toxin
21	1151.8	26.2	2710	19 AAM68387	Clostridium diffic
22	1151.4	26.2	3909	23 ABB43661	mRNA for hPac prot
23	1151.4	26.2	4590	23 AAB26790	Human pancreatic c
24	1151.4	26.2	4590	24 AAO26791	Human cadherin (CA
25	1151.4	26.2	4590	24 AAO26791	Human cadherin (CA
26	1151.4	26.2	4591	22 ABB22977	Novel human diagno
27	1150.9	26.2	2233	24 ABU01029	S. pneumoniae type
28	1150.9	26.2	2234	21 AAY81502	Streptococcus pneu
29	1148.8	26.1	3269	24 ABB54436	Papaya leaf-dictor
30	1147.6	26.1	15281	15 AAR44929	T. niyeum Cyclospo
31	1147.4	26.1	4081	22 ABB62958	Drosophila melanog
32	1147.3	26.1	3457	15 AAR62504	Large polypeptid
33	1142.5	26.0	3443	20 AAM64559	Polypeptid encode
34	1142.1	26.0	2681	24 ABB19025	Pathogen specific
35	1140.4	25.9	4796	22 ABB58665	Drosophila melanog
36	1138.9	25.9	4555	23 AAM52106	Rat fat 3 protein
37	1134.7	25.8	4599	24 ABB56837	Human LRP1B protei
38	1133.5	25.8	4472	22 ABB60101	Drosophila melanog
39	1133.3	25.7	3457	20 AAM64560	Polypeptid encode
40	1131.7	25.7	4636	22 AAE11937	Human CG168 (or CS
41	1130.2	25.7	3542	22 AAB62142	P. falciparum PCR3
42	1130.2	25.7	7718	22 ABB11811	Novel human diagno
43	1128.8	25.7	3097	22 ABB62967	Drosophila melanog
44	1127.1	25.6	4152	22 AAG64992	Shrimp white spot
45	1126.6	25.6	5635	23 ABB60991	Novel human protei

## ALIGNMENTS

RESULT 1	AAB48342	standard; Protein. 664 AA.
ID	AAB48342	
XX	AC	AAB48342;
XX	AC	20-APR-2001 (first entry)
XX	XX	S. pneumoniae Sp128 polypeptide.
XX	XX	Immunogenic; Sp128; Sp130; pneumococcal; otitis media; nasopharyngeal;
XX	KW	bronchial; lung; blood; infection; immune response; immunotherapy;
XX	KW	antibacterial; auditory; vaccine.
XX	XX	Streptococcus pneumoniae.
OS	OS	WO200076540-A2.
XX	XX	21-DEC-2000.
XX	PD	09-JUN-2000; 2000MO-US15925.
XX	XX	10-JUN-1999; 99US-0138453.
XX	XX	(MED1-) MED IMMUNE INC.
XX	XX	Adamou JE, Choi GH;
XX	XX	WPI; 2001-112197/12.
DR	DR	N-PSDB; AAC84741.
XX	XX	New vaccines comprising Sp128 or Sp130 polypeptides, for treating and
PT	PT	

PT preventing pneumococcal infections, particularly infections caused by  
PT Streptococcus, e.g. otitis media, nasopharyngeal, bronchial, lung or  
PT blood infections -  
XX  
PS Claim 4; Page 47-50; 54pp; English.  
XX  
CC The invention relates to novel immunogenic polypeptides, Spi28 and Spi30  
CC from S. pneumoniae. Vaccines comprising the polypeptides are useful for  
CC the treatment and prevention of pneumococcal infections, particularly  
CC infections caused by Streptococcus, such as otitis media, nasopharyngeal,  
CC bronchial, lung or blood infections. The antigens are used as immunogenic  
CC agents to stimulate an immune response. The antisera and antibodies may  
CC also be used in diagnosing and treating pneumococcal infections.  
CC Recombinant polypeptides serve as a mechanism for stimulating production  
CC of antibodies for use in passive immunotherapy, diagnostic reagents, and  
CC as reagents in other processes such as affinity chromatography. The  
CC present sequence represents the S. pneumoniae Spi28 polypeptide.  
XX  
SQ Sequence 664 AA;  
  
Query Match 100.0%; Score 4398; DB 22; Length 664;  
Best Local Similarity 100.0%; Pred. No. 3.7e-73;  
Matches 664; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 YPVVLADTSSSEDALNISDKVAENKEKHENIHSAMETSDQFKEKKTAVIKEKVVSKN 60  
DB 1 YPVVLADTSSSEDALNISDKVAENKEKHENIHSAMETSDQFKEKKTAVIKEKVVSKN 60  
  
QY 61 PVIDNNTSNEEAKIKEENSNSQGDYTDTSFVNKNTENPKKEDKVYVIAEFKDKESGEKAI 120  
DB 61 PVIDNNTSNEEAKIKEENSNSQGDYTDTSFVNKNTENPKKEDKVYVIAEFKDKESGEKAI 120  
  
QY 121 KELSLKNTKVLTYVDRIFNFGSAIETTPDNLDKIKQIEGISSVERAQKQVPMNHARKEI 180  
DB 121 KELSLKNTKVLTYVDRIFNFGSAIETTPDNLDKIKQIEGISSVERAQKQVPMNHARKEI 180  
  
QY 181 GVEEAIDYLSKNAPGKNFGRGMVISNIDGTDRYRKAMRIDDAKASMRFKKEDLKG 240  
DB 181 GVEEAIDYLSKNAPGKNFGRGMVISNIDGTDRYRKAMRIDDAKASMRFKKEDLKG 240  
  
QY 241 TDKNYWLSDKIPAFNYNYNGKITVEKYDDGRDYDPHGMHLAGLNDTQDKNFNG 300  
DB 241 TDKNYWLSDKIPAFNYNYNGKITVEKYDDGRDYDPHGMHLAGLNDTQDKNFNG 300  
  
QY 301 IDGIAPNAQIFSYKMYSDAGSGFAGDETWFHAIEDSIKHNVDVSVSSGFTGTGLVGEKY 360  
DB 301 IDGIAPNAQIFSYKMYSDAGSGFAGDETWFHAIEDSIKHNVDVSVSSGFTGTGLVGEKY 360  
  
QY 361 WQAIRALKRAGIPMVVATGNVATSNATSSSSDVLVANNHLKMTDTGNVTRTAHEDAIAVAS 420  
DB 361 WQAIRALKRAGIPMVVATGNVATSNATSSSSDVLVANNHLKMTDTGNVTRTAHEDAIAVAS 420  
  
QY 421 AKNQTVFEDKVNIGGESFYRNIGAFDFDKSKITTNEDGTGKPSKLVYIGKQDQDLIG 480  
DB 421 AKNQTVFEDKVNIGGESFYRNIGAFDFDKSKITTNEDGTGKPSKLVYIGKQDQDLIG 480  
  
QY 481 LDLRGKIAVMRIYTKDLKNAPFKKAMDKGARAIMVNTVNYNRDNTWELPAMGYEADBG 540  
DB 481 LDLRGKIAVMRIYTKDLKNAPFKKAMDKGARAIMVNTVNYNRDNTWELPAMGYEADBG 540  
  
QY 541 TKSQVFSISGDDGVKLWNNINPDKTEVKRNKEDFKDKLEQYYPIDMESFNNSPNVGD 600  
DB 541 TKSQVFSISGDDGVKLWNNINPDKTEVKRNKEDFKDKLEQYYPIDMESFNNSPNVGD 600  
  
QY 601 EKEIDFKFAPDTDKELYEDDIIVPAGSTSWGPRIDLKLPDVSAPGKNIKSTLNIVNGKS 660  
DB 601 EKEIDFKFAPDTDKELYEDDIIVPAGSTSWGPRIDLKLPDVSAPGKNIKSTLNIVNGKS 660  
  
QY 661 TYGY 664  
DB 661 TYGY 664

RESULT 2  
AAV81710  
ID AAV81710 standard; Protein; 2120 AA.  
XX  
AC AAV81710;  
XX  
DT 02-JUN-2000 (first entry)  
XX  
DE Streptococcus pneumoniae protein sequence ID3.  
XX  
KW Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;  
KW bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;  
KW kidney disease; diabetes; immunosuppressive disorder; otitis media;  
KW pneumococcal septicaemia; sinusitis; meningitis; therapy.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN WO200006738-A2.  
XX  
PD 10-FEB-2000.  
XX  
PF 27-JUL-1999; 99WO-GB02452.  
XX  
PR 27-JUL-1998; 98GB-0016336.  
XX  
PR 19-MAR-1999; 99US-0125329.  
XX  
PA (MICR-) MICROBIAL TECHNIQS LTD.  
XX  
PI Le Page RWF, Wells JM, Hanniffy SB, Hanabro PM;  
XX  
XX WPI; 2000-195301/17.  
DR N-PSDB; AAZ91806.  
XX  
XX Streptococcal proteins and polynucleotides useful for diagnosis,  
PT treatment and prophylaxis of bacterial infections -  
XX  
XX Claim 2; Page 41-42; 76pp; English.  
XX  
CC This sequence represents a Streptococcus pneumoniae protein of the  
CC invention. The proteins (or their homologues, derivatives and/or  
CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic  
CC compositions comprising the proteins are useful as vaccines and also in  
CC diagnostic assays. The sequences are useful for the detection or  
CC diagnosis of S. pneumoniae infection, by contacting a sample to be tested  
CC with them. Agents capable of antagonising, inhibiting or interfering with  
CC the function or expression of the protein or polypeptide are useful in  
CC medical compositions in the treatment or prophylaxis of S. pneumoniae  
CC infection. As the sequences can be used to treat S. pneumoniae infection,  
CC they can be used to treat bacterial pneumonia, which has high rates in  
CC young children, the elderly, and in patients with predisposing conditions  
CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,  
CC or with immunosuppressive disorders, especially AIDS. They can also be  
CC used to treat pneumococcal septicaemia, otitis media, sinusitis, and  
CC meningitis.  
XX  
SQ Sequence 2120 AA;  
  
Query Match 100.0%; Score 4398; DB 21; Length 2120;  
Best Local Similarity 100.0%; Pred. No. 3.3e-71;  
Matches 664; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 YPVVLADTSSSEDALNISDKVAENKEKHENIHSAMETSDQFKEKKTAVIKEKVVSKN 60  
DB 1 YPVVLADTSSSEDALNISDKVAENKEKHENIHSAMETSDQFKEKKTAVIKEKVVSKN 60  
  
QY 61 PVIDNNTSNEEAKIKEENSNSQGDYTDTSFVNKNTENPKKEDKVYVIAEFKDKESGEKAI 120  
DB 61 PVIDNNTSNEEAKIKEENSNSQGDYTDTSFVNKNTENPKKEDKVYVIAEFKDKESGEKAI 120  
  
QY 121 KELSLKNTKVLTYVDRIFNFGSAIETTPDNLDKIKQIEGISSVERAQKQVPMNHARKEI 180  
DB 121 KELSLKNTKVLTYVDRIFNFGSAIETTPDNLDKIKQIEGISSVERAQKQVPMNHARKEI 180

QY 181 GVEEAIDYLSKINAPFGKDPGRGMVINSIDTGTDRHAKARIDDDAKASNRFKKEDLKG 240  
DB 181 GVEEAIDYLSKINAPFGKDPGRGMVINSIDTGTDRHAKARIDDDAKASNRFKKEDLKG 240  
QY 241 TDKNWLSDKIPHANVYNGKITVEKYDDGRDYPDPHGMHAGIAGLNDTEODIKNPG 300  
DB 241 TDKNWLSDKIPHANVYNGKITVEKYDDGRDYPDPHGMHAGIAGLNDTEODIKNPG 300  
QY 301 IDGIAFNQIFSYKMYSDAGSGFAGDETFMFAIEDSIKHNDVVSVSFGFTGLVGEKY 360  
DB 301 IDGIAFNQIFSYKMYSDAGSGFAGDETFMFAIEDSIKHNDVVSVSFGFTGLVGEKY 360  
QY 361 WQAIRALKRAGIPMVVATGNVATSSASSSSWDLVANNHLMKMTDGTGVTRTAAHEDAIAVAS 420  
DB 361 WQAIRALKRAGIPMVVATGNVATSSASSSSWDLVANNHLMKMTDGTGVTRTAAHEDAIAVAS 420  
QY 421 AKNQVFEFDKVNIGESFKYRNIGAFDPKSKITTNEDEGTAKPSKLFYIGKQDQDLIG 480  
DB 421 AKNQVFEFDKVNIGESFKYRNIGAFDPKSKITTNEDEGTAKPSKLFYIGKQDQDLIG 480  
QY 481 LDLRGKIVMDRIYTKDLKNAFKKAMDGARAIMVNTVNYNRDNTMELPMAGYEADG 540  
DB 481 LDLRGKIVMDRIYTKDLKNAFKKAMDGARAIMVNTVNYNRDNTMELPMAGYEADG 540  
QY 541 TKSQVFSISGDDGVKLMNINPDKTEVYKRNKEDFKDLQEQYPIDMESEFNSKNPNVGD 600  
DB 541 TKSQVFSISGDDGVKLMNINPDKTEVYKRNKEDFKDLQEQYPIDMESEFNSKNPNVGD 600  
QY 601 EKEIDFKAPDPTDKELYKEDIIVPAGSTSWGFRIDLKLPDVSAPGKNIKSTLVYNGKS 660  
DB 601 EKEIDFKAPDPTDKELYKEDIIVPAGSTSWGFRIDLKLPDVSAPGKNIKSTLVYNGKS 660  
QY 661 TYGY 664  
DB 661 TYGY 664

## RESULT 3

ABU01020 standard; Protein; 2140 AA.

XX ABU01020;  
XX AC  
XX DT 11-FEB-2003 (first entry)  
XX DE S. pneumoniae type 4 strain protein from coding region #590.  
XX KW Bacterial meningitis; pneumonia; sepsis; otitis media;  
XX KW ear infection; antiinflammatory; antibacterial; immunostimulant;  
XX KW auditory; respiratory; gene therapy; vaccine.  
XX OS Streptococcus pneumoniae type 4 strain.  
XX PN WO200277021-A2.  
XX PD 03-OCT-2002.  
XX PF 27-MAR-2002; 2002WO-1B02163.  
XX PR 27-MAR-2001; 2001GB-0007658.  
XX PA (CHIR-) CHIRON SPA.  
XX PA (GENO-) INST GENOMIC RES.  
XX PI Masignani V, Tetteijn H, Frazer C;  
XX DR MPI; 2003-040579/03.  
XX DR N-PSDB; ABX06302.  
XX PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
XX PT useful as medicaments for treating or preventing a disease or infection  
XX PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media  
XX PT or ear infection

XX Claim 1; SEQ ID No 1180; 56bp; English.  
PS The invention relates to a protein comprising or having at least 50%  
XX identity to any of the 2469 amino acid sequences, identified in the  
CC specification (available on a computer readable format), or its fragment,  
CC expressed from 2469 of 2489 identified DNA coding regions from the  
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as  
CC AB856454. Also included are an antibody which binds one of the  
CC proteins, treating a patient by administering the protein, DNA or  
CC antibody (in a composition), a kit comprising first and second primers,  
CC which are the nucleic acid cited above or fragments between nucleotides  
CC 8-100 of a sequence not defined in the specification, for amplifying a  
CC target sequence contained within a Streptococcus nucleic acid sequence,  
CC where the first primer is substantially complementary to the target  
CC sequence and the second primer is substantially complementary to the  
CC complement of the target sequence, and where the parts of the primers  
CC having substantial complementarity define the termini of the target  
CC sequence to be amplified, assay comprising contacting a test compound  
CC with the protein, and determining whether the test compound binds to the  
CC protein and a Streptococcus pneumoniae bacterium, where one or more  
CC genes encoding the proteins has been rendered inactive. The proteins,  
CC nucleic acid molecules, antibody and compositions are useful as  
CC medicaments for treating or preventing a disease or infection due to  
CC streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,  
CC sepsis, otitis media or ear infection. They are also useful in developing  
CC vaccines, diagnostics and antibiotics. The methods are useful for  
CC identifying immunodominant proteins. The present sequence is one of  
CC the 2469 proteins expressed by the identified coding regions from the  
CC genomic sequence.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 2140 AA;

Query Match 99.6%; Score 4381.7; DB 24; Length 2140;

Best Local Similarity 99.6%; Pred. No. 76-71; 0; Indels 3; Gaps 2;

Matches 664; Conservative 0; Mismatches 0; Indels 3; Gaps 2;

QY 1 Y-P--VVLADTSSSEDLNISDEKEVAENKEKHENHSAMETSODPFKKTAIVKEKEVV 57  
DB 19 YAPNEVVLADTSSSEDLNISDEKEVAENKEKHENHSAMETSODPFKKTAIVKEKEVV 78  
QY 58 SKNPVIDNNTSNEBAKIKENSNSKSGDYTDSFVNKXTENPKEDKVVYIAEPKDSGE 117  
DB 79 SKNPVIDNNTSNEBAKIKENSNSKSGDYTDSFVNKXTENPKEDKVVYIAEPKDSGE 138  
QY 118 KAIKELSLKNTKVLVYDRIFNGSAIETTPDNLDKIKOIEGISVVERAOKVQPMNHAR 177  
DB 139 KAIKELSLKNTKVLVYDRIFNGSAIETTPDNLDKIKOIEGISVVERAOKVQPMNHAR 198  
QY 178 KEIGVEEAIDYLSKINAPFGKDPGRGMVINSIDTGTDRHAKARIDDDAKASNRFKKED 237  
DB 199 KEIGVEEAIDYLSKINAPFGKDPGRGMVINSIDTGTDRHAKARIDDDAKASNRFKKED 258  
QY 238 LKGTDKNWLSDKIPHANVYNGKITVEKYDDGRDYPDPHGMHAGIAGLNDTEODIKN 297  
DB 259 LKGTDKNWLSDKIPHANVYNGKITVEKYDDGRDYPDPHGMHAGIAGLNDTEODIKN 318  
QY 298 FNGIDGIAFNQIFSYKMYSDAGSGFAGDETFMFAIEDSIKHNDVVSVSFGFTGLVYG 357  
DB 319 FNGIDGIAFNQIFSYKMYSDAGSGFAGDETFMFAIEDSIKHNDVVSVSFGFTGLVYG 378  
QY 358 EKYWQAIRALKRAGIPMVVATGNVATSSASSSSWDLVANNHLMKMTDGTGVTRTAAHEDAIA 417  
DB 379 EKYWQAIRALKRAGIPMVVATGNVATSSASSSSWDLVANNHLMKMTDGTGVTRTAAHEDAIA 438  
QY 418 VASAKNQVFEFDKVNIGESFKYRNIGAFDPKSKITTNEDEGTAKPSKLFYIGKQDQD 477  
DB 439 VASAKNQVFEFDKVNIGESFKYRNIGAFDPKSKITTNEDEGTAKPSKLFYIGKQDQD 498

QY 478 LIGDLRGKIAVMRIYTKDLKNAFKKAMDKGARAIMVNTVYNNRDNWTELPAWGVEA 537  
 |||||  
 DB 499 LIGDLRGKIAVMRIYTKDLKNAFKKAMDKGARAIMVNTVYNNRDNWTELPAWGVEA 558  
 |||||  
 QY 538 DEGTQSQVFSISGDDGVKLWNNINPDKTEVRRNKEDFKDLQYYPIDMESFNSKNPN 597  
 |||||  
 DB 559 DEGTQSQVFSISGDDGVKLWNNINPDKTEVRRNKEDFKDLQYYPIDMESFNSKNPN 618  
 |||||  
 QY 598 VGDEKEIDFKAPDPTDKELYKEDIIVPAGSTGWSGPRIDLLLPDVSAFGKNIKSTLNVIN 657  
 |||||  
 DB 619 VGDEKEIDFKAPDPTDKELYKEDIIVPAGSTGWSGPRIDLLLPDVSAFGKNIKSTLNVIN 678  
 |||||  
 QY 658 GKSTYGY 664  
 |||||  
 DB 679 GKSTYGY 685

## RESULT 4

AAW61246  
 ID AAW61246 standard; Protein; 1007 AA.

XX  
 AC AAW61246;

DT 02-OCT-1998 (first entry)

XX Streptococcus pneumoniae SPI22 protein.

DE Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;  
 KW detection; pneumonia; otitis media; meningitis.

XX Streptococcus pneumoniae.

XX WO9818930-A2.

XX 07-MAY-1998.

PF 30-OCT-1997; 97WO-US19422.

PR 31-OCT-1996; 96US-0029960.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Choi GH, Hromockyj A, Johnson LS, Kunsch CA;

XX WPI; 1998-272224/24.

DR N-PSDB; AAV27431.

XX Nucleic acid encoding antigenic peptide(s) from Streptococcus  
 PT pneumoniae - or their epitope-containing fragments, useful in  
 PT protective or therapeutic vaccines, and for diagnosis

PS Claim 11; Page 92-93; 118pp; English.

XX The present sequence represents a protein from Streptococcus pneumoniae.  
 CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein  
 CC can be useful in vaccines for inducing protective antibodies against  
 CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.  
 CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid  
 CC are used to detect Streptococcus infection (by usual hybridisation or  
 CC amplification methods), also for isolating Streptococcus genes or their  
 CC allelic variants. The protein can be used similarly to detect specific  
 CC antibodies in standard immunoassays, especially for diagnosing or  
 CC monitoring infections. Antibodies which bind the protein are used to  
 CC detect corresponding antigens, to purify the protein and for passive  
 CC immunisation (optionally coupled to a toxin). Vaccines are administered,  
 CC e.g. by injection, orally or through the skin, typically at 0.01-1000  
 CC (especially 10-300) mu g/ml per dose.

XX Sequence 1007 AA;

Query Match 94.8%; Score 4171; DB 19; Length 1007;  
 Best Local Similarity 100.0%; Pred. No. 4, 4e-68;  
 Matches 627; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 ETSODFKKKTAVTIKEKEVVSKNPVIDNNTSNEAKIKEENSNSKSGDYTDSFVNKNNTEN 97  
 |||||  
 DB 1 ETSODFKKKTAVTIKEKEVVSKNPVIDNNTSNEAKIKEENSNSKSGDYTDSFVNKNNTEN 60  
 |||||  
 QY 98 PKKEDKVYVIAEFKDKESGEKAIKELSSLKNTKVLTYDYRIFNGSAIETTDPNLDKI KOI 157  
 |||||  
 DB 61 PKKEDKVYVIAEFKDKESGEKAIKELSSLKNTKVLTYDYRIFNGSAIETTDPNLDKI KOI 120  
 |||||  
 QY 158 EGISVSERAQKVQPMNHNHARKEIGVEEADYKLSINAPFGKNFGGRGMVINSIDTGTDYR 217  
 |||||  
 DB 121 EGISVSERAQKVQPMNHNHARKEIGVEEADYKLSINAPFGKNFGGRGMVINSIDTGTDYR 180  
 |||||  
 QY 218 HKAMRIIDDDAKASMRFKKEDLKGTDKNWLSDKIPHAFNYNGGKITVKEYDDGDRDYDP 277  
 |||||  
 DB 181 HKAMRIIDDDAKASMRFKKEDLKGTDKNWLSDKIPHAFNYNGGKITVKEYDDGDRDYDP 240  
 |||||  
 QY 278 HGMHIAGILAGNTEQDIKNFNGIDGIAPNAQIFSYKMSDAGSGFAGDETMFHAIEDSI 337  
 |||||  
 DB 241 HGMHIAGILAGNTEQDIKNFNGIDGIAPNAQIFSYKMSDAGSGFAGDETMFHAIEDSI 300  
 |||||  
 QY 338 KHNVDVSVSGFTGTGLVGBKYNQAIIRALRKAGIPMVVATGNVATSSSSSWDLVANNH 397  
 |||||  
 DB 301 KHNVDVSVSGFTGTGLVGBKYNQAIIRALRKAGIPMVVATGNVATSSSSSWDLVANNH 360  
 |||||  
 QY 398 LKMTDTGNVTRTAAHEDAIASAKNQTVEFDKYNIGGESPKYRNIGAFFDKSKITTNED 457  
 |||||  
 DB 361 LKMTDTGNVTRTAAHEDAIASAKNQTVEFDKYNIGGESPKYRNIGAFFDKSKITTNED 420  
 |||||  
 QY 458 GTKAPSKLKFVYIGKGQDGLIGDLRGKIAVMRIYTKDLKNAFKKAMDKGARAIMVYN 517  
 |||||  
 DB 421 GTKAPSKLKFVYIGKGQDGLIGDLRGKIAVMRIYTKDLKNAFKKAMDKGARAIMVYN 480  
 |||||  
 QY 518 TVNYNRDNWTELPAWGVEADEGTQSQVFSISGDDGVKLWNNINPDKTEVRRNKEDFK 577  
 |||||  
 DB 481 TVNYNRDNWTELPAWGVEADEGTQSQVFSISGDDGVKLWNNINPDKTEVRRNKEDFK 540  
 |||||  
 QY 578 DKLEQYYPIDMESFNSKNPNVDEKEIDFKAPDPTDKELYKEDIIVPAGSTGWSGPRIDLL 637  
 |||||  
 DB 541 DKLEQYYPIDMESFNSKNPNVDEKEIDFKAPDPTDKELYKEDIIVPAGSTGWSGPRIDLL 600  
 |||||  
 QY 638 LKPDVSAFGKNIKSTLNVIN GKSTYGY 664  
 |||||  
 DB 601 LKPDVSAFGKNIKSTLNVIN GKSTYGY 627

## RESULT 5

ABP54664

ID ABP54664 standard; Protein; 1007 AA.

XX  
 AC ABP54664;

XX 04-SEP-2002 (first entry)

XX S. pneumoniae SPI22 protein sequence SEQ ID NO:216.

XX Streptococcus pneumoniae; epitope; vaccine; antigenic protein;  
 KW antibacterial; Streptococcal infection; detection.

XX Streptococcus pneumoniae.

XX US2002061545-A1.

XX 23-MAY-2002.

XX 22-JAN-2001; 2001US-0765272.

XX 30-OCT-1997; 97US-0961083.

XX (CHOI/) CHOI G H.

XX (KUNS/) KUNSCH C A.

XX (BARA/) BARASH S C.

XX (DILL/) DILLON P J.

PA (DOUG/) DOUGHERTY B.  
 PA (FANN/) FANNON M R.  
 PA (ROSE/) ROSEN C A.  
 XX  
 XX  
 PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR,  
 PI Rosen CA;  
 DR WPI, 2002-479261/51.  
 DR N-PSDB; ABQ84899.  
 XX  
 PT New Streptococcus pneumoniae antigens, useful for detecting  
 PT Streptococcus and for preventing or attenuating disease caused by  
 PT Streptococcus infection -  
 XX  
 PS Claim 11, Page 50; 70pp; English.  
 XX  
 CC ABQ84792 to ABQ84904 represents nucleic acids which encode the  
 CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669.  
 CC The S. pneumoniae antigens have antibacterial activity and can be  
 CC used in vaccines. The S. pneumoniae antigens can also be used to  
 CC prevent or attenuate a Streptococcal infection in an animal. The  
 CC polynucleotides encoding the S. pneumoniae antigens can be used to  
 CC detect Streptococcus nucleic acids. ABQ84905 to ABQ85130 represent  
 CC primers used in the cloning of S. pneumoniae ORFs (open reading frames)  
 CC which are used in an example from the present invention.  
 CC  
 SQ Sequence 1007 AA;

Query Match 94.8%; Score 4171; DB 23; Length 1007;  
 Best Local Similarity 100.0%; Pred. No. 4,4e-68;  
 Matches 627; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 38 ETSOPFKEKTAIVIEKEVNSKPIVINDNTSNEAKIKESNKSOGYDTSFVNKNEN 97  
 DB 1 ETSOPFKEKTAIVIEKEVNSKPIVINDNTSNEAKIKESNKSOGYDTSFVNKNEN 60  
 OY 98 PKKEDKVVYIAEFKDKESGEKAIKELSLKNTKVLTYDRIFNGSAIETTPNDLKIKOI 157  
 DB 61 PKKEDKVVYIAEFKDKESGEKAIKELSLKNTKVLTYDRIFNGSAIETTPNDLKIKOI 120  
 OY 158 EGISSEVERAOKVQPMNTHARKEIGVEAIDYKKSINAPFGKNFDRGMVINSIDTGYR 217  
 DB 121 EGISSEVERAOKVQPMNTHARKEIGVEAIDYKKSINAPFGKNFDRGMVINSIDTGYR 180  
 OY 218 HKAMIDDDAKASMKFKKEDLKTGDKNWLSPKIPHAFYVNGKITVEKYDGDGDP 277  
 DB 181 HKAMIDDDAKASMKFKKEDLKTGDKNWLSPKIPHAFYVNGKITVEKYDGDGDP 240  
 OY 278 HGMIAGILAGNDTEODIKNFNGIDGIAPNAOIFSYKNYSDAGSGFAGDETFFHAIEDI 337  
 DB 241 HGMIAGILAGNDTEODIKNFNGIDGIAPNAOIFSYKNYSDAGSGFAGDETFFHAIEDI 300  
 OY 338 KHNVVSVSSGFTGTGLVGEKXWQAIRLRKAGIPMYVATGNVATSSASSSMDLVANNH 397  
 DB 301 KHNVVSVSSGFTGTGLVGEKXWQAIRLRKAGIPMYVATGNVATSSASSSMDLVANNH 360  
 OY 398 LKMDTGTGVTTRAHAEADAIAVASAKQVPEPKNVIIGESFYNIGAFPDKSKITTTED 457  
 DB 361 LKMDTGTGVTTRAHAEADAIAVASAKQVPEPKNVIIGESFYNIGAFPDKSKITTTED 420  
 OY 458 GTKAPSKLKFVYIGKQODODLIGDLRGKIAMDRITTKDLKNAFKKMDKARAIMVYN 517  
 DB 421 GTKAPSKLKFVYIGKQODODLIGDLRGKIAMDRITTKDLKNAFKKMDKARAIMVYN 480  
 OY 518 TVNYNRDNWTELPMAGYEADGRTKSQVFSISGDDGVKLMNINDEKKTVEVRNNKEDPK 577  
 DB 481 TVNYNRDNWTELPMAGYEADGRTKSQVFSISGDDGVKLMNINDEKKTVEVRNNKEDPK 540  
 OY 578 DKLGEYVPIIDMESFSNKNPNVGDKEIDEKPAFDPDKLKYEDITVPAGSTSMGRIDLL 637  
 DB 541 DKLGEYVPIIDMESFSNKNPNVGDKEIDEKPAFDPDKLKYEDITVPAGSTSMGRIDLL 600  
 OY 638 LKPDVSABGKNIKSTLVNJKSTGYG 664

DB 601 LKPDVSABGKNIKSTLVNJKSTGYG 627  
 |||||  
 RESULT 6  
 AAM68388  
 ID AAM68388 standard; Protein; 2366 AA.  
 XX  
 XX AAM68388;  
 AC  
 DT 07-DEC-1998 (first entry)  
 XX  
 DE Clostridium difficile toxin B.  
 XX  
 XX Antitoxin; vaccine; cytotoxin; toxin B; intoxication; immunogen;  
 KM pseudomembranous enterocolitis.  
 XX  
 XX Clostridium difficile.  
 XX  
 OS WO9808540-A1.  
 XX  
 PN 05-MAR-1998.  
 PD  
 XX 28-AUG-1997; 97WO-US15394.  
 XX  
 PR 28-AUG-1996; 96US-0704159.  
 XX  
 PA (OPHI-) OPHIDIAN PHARM INC.  
 PI Thallely BS, Williams JA;  
 XX  
 DR WPI: 1998-230234/20.  
 DR N-PSDB; AAV30561.  
 XX  
 PT Host cell containing recombinant expression vector encoding  
 PT Clostridium botulinum type B or E toxin - useful to treat humans  
 PT and other animals at risk of intoxication with clostridial toxin  
 XX  
 PS Example 18; Page 241-249; 428pp; English.  
 XX  
 CC This is the amino acid sequence of Clostridium difficile toxin B,  
 CC deduced from the coding region (see AAV30561) of the toxin B gene.  
 CC Fragments of the toxin B gene have been cloned into various  
 CC prokaryotic expression systems, and assessed for the ability to  
 CC express recombinant toxin B protein in E. coli. It would be  
 CC advantageous to use simple and inexpensive prokaryotic expression  
 CC systems to produce and purify high levels of recombinant toxin B  
 CC for immunisation purposes. The invention specifically relates to  
 CC recombinant proteins derived from Clostridium botulinum toxins  
 CC (see AAM68389-400) and their use as immunogens for the production of  
 CC vaccines and antitoxins.  
 CC  
 SQ Sequence 2366 AA;  
 XX  
 Query Match 27.0%; Score 1189.3; DB 19; Length 2366;  
 Best Local Similarity 12.4%; Pred. No. 3.8e-09;  
 Matches 292; Conservative 143; Mismatches 198; Indels 1717; Gaps 119;  
 OY 1 YP-----VV-----LADTSSSEDALNISDEKVAENK----- 27  
 DB 34 YHNNSENVTEVEKYLKLDINDSLDIY-IDTYKSGGRKALKKFKELYLTVEVLKNNNLT 92  
 OY 28 --EKHENIH----- 36  
 DB 93 PVKE--NLHFWIGQINDTAINYINOMKDVNSDVNVVFPYDSNAFLINTLTKTVESAI 150  
 OY 37 METSODFE-----KKTAVI--KEKEV-----SKNP--VIDNT----- 67  
 DB 151 NDTLESFRENLDNDRFDYNNKFFRKQMOIYDKQNFINYKAQRENEBELIIDIIVKTYL 210  
 OY 68 SNEBAK-IKEENS-----NK-SQ----- 83  
 DB 211 SNEYSKIDEIDELNTYIESLNKNTITQNSGNDVNRNPEEFKNGSSFNLYEQBELVERMNLAAASD 270

QY 84 -----CDYTD----- 88  
 Db 271 ILRISALKEIGMYLDVDMPLGOIOPDLFESIEKPSVTVDFWMTKLEAIMKYEYIPEY 330  
 QY 89 -----SF-----VN----- 92  
 Db 331 TSEHDMLEEVQSSFEFSLASKSDKSEIFSSLDMEASPLEVKIAPNSKGIINOGLISV 390  
 QY 93 -----KXNENPK-KED----- 102  
 Db 391 KDSYCSNLIVKQIENRYKILNLSNPAISEDNDFTNTTTFIDSIAMAEANADNGRFMMEL 450  
 QY 103 -KVVVIAEFKKE-----SGEKA-----IKELS-----SLKNTKV----- 131  
 Db 451 KYLRVGFPPDKTTINLSGPEAYAAAYQDLLMFKEGSMNTHLEADLRNFEISKNTISQ 510  
 QY 132 -----LYTYD-----RI-PNGSAIETTPDNLDKIKOI-----EGISS 162  
 Db 511 STEQEMASLWSPDDARAKAQPEYKRNYPEGSLGE--DDNLDPSQNIIVVDKEYLLEKISS 568  
 QY 163 VERA----- 166  
 Db 569 LARSSERYIHIVQLQGDKISYEAACNLFAKTPYDSVLFOKNIEDSEIAYVYNNPGDGEI 628  
 QY 167 OKV-----OPM-----MNHARKE-----IG-----VEEADYLK----- 190  
 Db 629 QEIDKYKIPSIIDRPKIKLTFIGHGKDFNTDIFAGFDVDSLSLSTEIAAADLAKEDISP 688  
 QY 191 -----SINA-----PFGK-----N 199  
 Db 689 KSIEINLLCCNMFYSINVEETYP-GKLLLVKDKISELMPISODSIIVSANQYEVRLN 747  
 QY 200 FDGR-----G----- 204  
 Db 748 SEGRELLDHSGEWINKESIIKDISKEYISFNPKENKITVKSKNLPSELSTLLQEIRNN 807  
 QY 205 -----MVISNDTGT----- 214  
 Db 808 SNSDIELEBEKVMLTECEINVSINSDTQIVEERIEBAKNLTSDSINYIKDEPKLIESISD 867  
 QY 215 ---DVR-----HKAMRID---DDAKASMF-KKE----- 236  
 Db 868 ALCDLKQONELEDHSFISPEIDSETDEGFSIRFINKETGESIFVETEKTIPEYANHITE 927  
 QY 237 ---DLKGT---DK-NYWLSDKI---PHAFNY-----YNGK----- 262  
 Db 928 EISKIKGTIFTVNGKLVKKNVLDTTHEVNTLNAAFFIQSLIEYNSKESLSNLSVAMKV 987  
 QY 263 -----IT-----VE---KYDDGRDYDP---HGMH-----IAGILAGNDE 292  
 Db 988 QVYAQLFSTGLNTIDAARKVWELVSTALDETIDL-PTLSEGLPIATIIDGVSGLAAIK 1046  
 QY 293 -----ODIKNFNGI-----DGIA-----P--- 306  
 Db 1047 ELSETSDPLROEIEAKIGIMAVNLTTATTITSLGSLGASFSILLVPLAGISAGISL 1106  
 QY 307 -NAQ-----IFSY-----KMYSD----- 318  
 Db 1107 VNNELVLRDKATKVVDYFKHVSIVETEGVFTLLDDKIMMPQDVLVISEIDFNNSIVLGK 1166  
 QY 319 -----AGSGFAGDETMFAIEDSIKH-----NVDVVS----- 346  
 Db 1167 CEIWRMEGGSG-----HTVTDDIDHFSAPSITYREPHLSIYDVLEVOKEELDLSKD 1218  
 QY 347 -----SSGFT-----CTGLV-----GEKWO----- 362  
 Db 1219 LMVLPNAPNRVFAWETGTPGLRSLENDGTLLDRIDNYEGEFYRYFAPTADALITTL 1278  
 QY 363 -----AIR-----ALRKAGIPMVVAT-----GNVATSAS----- 386  
 Db 1279 KPREDYTNIRINDSNTRSFIVP-IITTEYIREKLSYFPGSGGTVALSQQVNMGINIE 1337

QY 387 ---SSSW-----DLV----- 393  
 Db 1338 LSESDVMIIDVNVVRDVTIESDKIKKGDLLIEGLISTLSIENKIILNSHEINFSGEVNG 1397  
 QY 394 AN-----NHLK----- 399  
 Db 1398 SNGFVSLTSLILEGINAIIEVDLLSKSYKLLISGELKILMNSNHIQOKIDYIGFNSQLQ 1457  
 QY 400 -----MTD----- 402  
 Db 1458 KNIPYSFVDSGKENGFGINGSTKEGLFVSELPDVLISKVYMDDSKPSFGYVSNLKDVK 1517  
 QY 403 -----TG-----NVRTAAHEDAIASASA---KNO----- 424  
 Db 1518 VITKDNVILTYLKKODIKISLSLTLODEKTIKLSNVHLDSEGVAEILKPMNRKGTNT 1577  
 QY 425 -----TVEF----- 428  
 Db 1578 SDLSMFLSMNIKSIFVNFQSNIKPILDANFIISGTTISIGQFEFICDENDNIOPYFIK 1637  
 QY 429 ----- 428  
 Db 1638 FNTLETNYLVGNRQNMIVEPNYDLDSDISSTVINFSQKLYGIDSCVKNVVISPN 1697  
 QY 429 --DKVNI-----GGESF----- 438  
 Db 1698 YTDENITPYETNTYPEVIVLDANYINEKINVININDLSIRYVNSDNGDNFILMSTSEE 1757  
 QY 439 ---KYRNTGAFDKS---KITTN-EDGTPAKPSK---LKFV---YIGKGQDQDLIGLDL 483  
 Db 1758 NKVSQVKIRFVNFVKDITLANKLSFNSDKQDPVSEIILSFTPSYEDG---LIGYDL 1813  
 QY 484 -----RGIAMVDRIY-----TKDLKNAFKKAMDK---GA 510  
 Db 1814 GLVSLYNEKFYNNFMMVSGLIYINDSLYFYKPPVNNLITGFTVVGDDKYFNPINGGA 1873  
 QY 511 RAI--MVNTVNY--NR----- 524  
 Db 1874 ASIGETIIDDKNYFNSQSVLQTVGFSTEDGFKYFAPANTLDENLEGEAIDFTGKLIIDE 1933  
 QY 525 ---DN-----WTEL-----PAMG----- 534  
 Db 1934 NIYFDDNYRGAVEMKELDGMHYFSPETGKAFGLAQIGDYKYFYNSDGMVQKGFVSIN 1993  
 QY 535 ---YEADGFK---SQV-----FSISG-----DDGVLKNNINPDKTE--- 567  
 Db 1994 DNKHYFDDSGVMKVGYTEIDGKHFYFAENGEMQIGVFNTEDGFKYFAHHNEDLGNEGEE 2053  
 QY 568 ---VKRNNK-----ED-----F-KDKLE-----OYPTID 587  
 Db 2054 ISYSGILNPNKIYFDDSFSTAVVGWKKLEDDGSKYFYFDEDTAEAYIGLSLNDGGYYFND 2113  
 QY 588 ---MES----- 591  
 Db 2114 DGIMQVFTINDKVYFSDSGIIESGVQINDDNVFYDDNGIYVQIGVFTSDGKYFAP 2173  
 QY 592 -NSNKN-----VG-----D-EKEID-FKAPDT----- 612  
 Db 2174 ANTVDNITYGOAVEYSGLVYRGEDVYFGETYTIETGTWYIDMENESDKYFNPETKACK 2233  
 QY 613 -----DKELY-KEDI--- 622  
 Db 2234 GINLIDDIKYFDEKIMRTGLISFENNYYFNENGEMQFYGINTEDKMFFGEGDGVQOI 2293  
 QY 623 ---VPAGS-----TSWGPRI DL-----LLKPD---VSAPGNKIIST 652  
 Db 2294 GVENPTDGFKYFAHQNTLDENFEGESINVTGW---LDLDEKRYFYFTDEVIATGTVI--- 2347  
 QY 653 LNVINGKSTY 662  
 Db 2348 ---IDGEBYY 2354



```

RESULT 7
AAR95011
ID AAR95011 standard: Protein; 2366 AA.
AC AAR95011;
XX
XX
DT 08-JUL-1996 (first entry)
XX
XX C. difficile toxin B.
DE
XX Toxin B; cytotoxin; enterotoxin; fusion protein; antitoxin;
KM diarrhoea; therapy; diagnosis; vaccine.
XX
OS Clostridium difficile VPI strain 10463 (ATCC 10463).
XX
XX MO9612802-A1.
XX
XX 02-MAY-1996.
XX
XX 23-OCT-1995; 95MO-US13737.
XX
XX 07-JUN-1995; 95US-0480604.
XX 24-OCT-1994; 94US-0329154.
XX 16-MAR-1995; 95US-0405496.
XX 14-APR-1995; 95US-0422711.
XX
XX (OPHI-) OPHIDIAN PHARM INC.
XX
XX Fitca JR, Kink JA, Padhye NV, Stafford DC, Thalley BS;
P1 Williams JA;
XX
XX WPI, 1996-230603/23.
DR N-PSDB; AAT29247.
XX
XX Fusion proteins comprising non-toxin protein and part of toxin
PT useful to form anti-toxins against Clostridium botulinum type A, and
PT C. difficile type toxins, and to treat C. difficile intoxication,
PT partic. diarrhoea
XX
XX Claim 36; Page 313-323; 434pp; English.
XX
XX Clostridium difficile VPI strain 10463 toxin B (AAR95011), the
CC product of the toxin B gene (AAT29247), is a cytotoxin associated
CC with diarrhoeic disease. It can be obtd. by expression in
CC transformed E. coli hosts of portions of DNA that together cover the
CC entire toxin B gene, toxin B, and portions of it (see also AAR95012-13,
CC AAR95311-72 and AAR95018), pref. expressed as fusions to polypeptide
CC affinity tags or maltose binding protein, are used to raise avian
CC antibodies useful as antitoxins or diagnostics, and in vaccine prodn.
CC
XX
SQ Sequence 2366 AA;
Query Match 27.0%; Score 1188.3; DB 17; Length 2366;
Beet Local Similarity 12.4%; Pred. No. 4e-09;
Matches 292; Conservative 142; Mismatches 199; Indels 1717; Gaps 119;
QY 1 YP-----VV-----LADTSSESDALNISKKEVAENK-----SA 27
DB 34 YHNSSENTVVEKYLKDKINSLTDIY-IDTYKKSGRNALKKKFKEYLVEVLELKNNNLT 92
QY 28 --EKENHNI-----KKTAVI--KEKEV-----SKNP--VIDNNT----- 36
DB 93 PVEK--NIHFVWIGQINDTAIYNQMKDVNSDYNVNFYDSNAFLINTLKTIVESA 150
QY 37 METSQDFKE-----KKTAVI--KEKEV-----SKNP--VIDNNT----- 67
DB 151 NDTLESFRENLDPRFDYNNKFFRKMEIIFYDKQKFINYYKQRENFELIIDIVKTYL 210
QY 68 SNEEK-IKENS-----NK-SQ----- 83
DB 211 SNEYSKEIDELNLTIESLNKITONSQNDVNRNFEFFKGESEFNLYEOELVERMNLAA 270
QY 84 -----GDYTD----- 88

```

```

DB 271 ILRISALKEIGWYLDVDMLPGIQDLPESIEKPSVTVDPEWMTKLEAIKMYEYIP 330
QY 89 -----SF-----VN----- 92
DB 331 TSEHFDMLDEBVOSSFESVLASKSDKSEIFSSLDGMEASPLEVXIAFNKGIINOGLIS 390
QY 93 -----KNTENPK-KED----- 102
DB 391 KDSYCSNLIYQIENRKYILNLSLNPALISEDNDNTTNTTFIDSIMAANANGRFMMEL 450
QY 103 -KVVYIAEFKDE-----SGEKA-----IKELS-----SLKNTKV----- 131
DB 451 KGYLRVGFPPDVKTITNLSCGEAYAAAYQDPLMFKESSMNIHLIADLRNFEISKTNISQ 510
QY 132 -----LYTD-----R-FCNSALETTPDNLDKIQI-----EGISS 162
DB 511 STEQEMASLWSFDDAKAKOPEEYKRNVPFGSLGE--DNDLDFSONIVDPKYLEKISS 568
QY 163 VERA----- 166
DB 569 LARSSERCYIHYIVQLOGDKISTEACNLFAKTPYDSVLPQKNIEDSEIAYVPGDGEI 628
QY 167 QKV-----QPM-----NMHARKE-----IG-----VEEAI DYK----- 190
DB 629 QEIDKXKIPSIISDRPKIKLFFIGHGKDEFPNTDIFAGDVSLSIETIAAIDLAKEDISP 688
QY 191 -----SINA-----PFGK-----N 199
DB 689 KSIEINLGCNMFPSYINVEETYP-GKLLLVKDKKISLMPSSISQSIIVSANQYEVARIN 747
QY 200 FDGR-----G----- 204
DB 748 SEGRBELDHSGEWINKESIIKDISKEYISFNPKENKITVSKNLPBLSTLLOEIRNN 807
QY 205 -----WVISNIDTGT----- 214
DB 808 SNSSDIELEKVMLTECEINVISNIDQIVBERIEAKNLSDSINIKDFKLEISID 867
QY 215 --DYR-----HKAMRID--DDAKASNR-FKKE----- 236
DB 868 ALCKOQNELEDSDHFSFEDISSETDECFISIRFYNKETGESIFVETEKITFSEYANHTE 927
QY 237 --DLKGT--DK--HYWLSDKI--PHAFN-----YNGK----- 262
DB 928 EISKIGTIFQYVNGKLVKKVNLDTTREVNLNAAFIQSLIEYNSKESLSNVAMKV 987
QY 263 -----IT-----VE-----KYDGRDYFDP--HGMD-----IAGILAGNDTE 292
DB 988 QVAQOLFSTGINTITTDAAKVVELVSTALDERIDLL-PTLSGCLPIATIIDGVSLGAIR 1046
QY 293 -----QDIKNFNGI-----DGIA-----P- 306
DB 1047 ELSETSPDLLQOEIEAKIGIYAVNLTTATTAIITSSLGISGFSILLPLAGISAGIPSL 1106
QY 307 -NAQ-----ISY-----KMSYD----- 318
DB 1107 VNNELVLDRKATKYVDYFKHVSIVETEGVFLLLDCKIMPODDIVISEIDFNNSIYLGK 1166
QY 319 -----AGSGFADETMFHAIEDSIKH-----NVDVVSV----- 346
DB 1167 CEIWRMEGSG-----HTVTDIDHFPESAPSIYREPLSLIYDULEVQKEELDLSKD 1218
QY 347 -----SGGFT-----GTGLV-----GEXYKQ----- 362
DB 1219 LMTVLPAENRVAFWETGWTPLRLSLENDGTKLDRIRDNVEGEPYRFAFIADALITTL 1278
QY 363 -----AIR-----AIRKAGIIPNVAT-----GNVATSA----- 386
DB 1279 KPRYEDTNIRINDSNTSRSPFIV-IITTEYIREKLSYSPYSGGTYALSIOYNNGINIE 1337
QY 387 ---SSSW-----DLV----- 393

```

Db 1338 LSESDVWIDVNVVDVTIESDKIKGDLIBGLSTLSIERNKIILNSHEINPFSGEVNG 1397  
 Qy 394 AN-----NHLK----- 399  
 Db 1398 SNGFVSLTFSILEGINAIIIEVDLLSKYKLLISGELKILMLNSNHIOOKIDYIGFNSQL 1457  
 Qy 400 -----MTD----- 402  
 Db 1458 KNIPYSVDSEKENGFGSTKEGLFVSELDDVVLISKVYMDSKSPFGYYSNNLKDVK 1517  
 Qy 403 -----TG-----NVRTAAHEDAIAVASA---KNQ----- 424  
 Db 1518 VITKDNVAILTGYYLKDDIKISLSLTLODEKTIKLSNVHLDSEGVAILKFWNRKGTNT 1577  
 Qy 425 -----TVFF----- 428  
 Db 1578 SDLSMFLESNMKISIFVNFLOSNIKFLDANFIISGTTISGQFEFICDENDNIQYFIK 1637  
 Qy 429 ----- 428  
 Db 1638 FNTLETNYLYVGNRQNMIVENPYDLDSDGISSTVINPFSQYLYGIDSCVKNKVISPNI 1697  
 Qy 429 --DKVNI-----GGESE----- 438  
 Db 1698 YTDEINITPVYETNNTYPEVIVLDANYINEKINVINDLIRYVWSNDGNDFILMSTSEE 1757  
 Qy 439 -----KYRNIGAPDKS---KITFN-EDGTWAPSK---LKFV---YIGKGQDDLIGLDL 483  
 Db 1758 NKVSQVKIRFVNVFKDKTLANKLSFNSFSDQVPVSEIILSFTPSYEDG---LIGYDL 1813  
 Qy 484 -----RGKIAVMRIY---TKDLKNAFKAMDK-----GA 510  
 Db 1814 GLVSLYNEKFYINNGMVMGSLIYINDSLYKPPVNNLITGVTVGDDKYFNPINGGA 1873  
 Qy 511 RAJ--MVNVTNYY-NR----- 524  
 Db 1874 ASIGETIIDDKNYNNQSCVLOTGVFSTEDGPKYFAPANTLDENLEGEAIDFTGKLIIDE 1933  
 Qy 525 -----DN-----PAMG----- 534  
 Db 1934 NIYPFDDNVGAVENKELDGMHYPSPETGKAFKGLNQIGDYKYFYFNSDGMQGFVSIN 1993  
 Qy 535 ---YEABGTG---SQV-----FSISG-----DQVKLWNNINPKKTE--- 567  
 Db 1994 DNKHYPDDSGVMKVGYTEIDGKHFFAENGEMQIGVFNTEDGKYFAHNNEDLGNEEGEE 2053  
 Qy 568 -----VKRNKK-----ED---F-KDKLE-----QYYPID 587  
 Db 2054 ISYSGILNFNNKIYYFDSDFTAVVGWKOLEDGSKYYFDEDTAEAYIGLSLINDGQYFND 2113  
 Qy 588 -----MBS-----P--- 591  
 Db 2114 DGMQGVFTINDKVFYFSDSGIIESGVQINDNYFYIDNDGIVQGVFDTSDGKYKYPAP 2173  
 Qy 592 -NSNKNP-----VC-----D-EKID-PKFAPDN----- 612  
 Db 2174 ANTVDNIVQGAVEYSGLVRVGEDVYFGETYTIETGTWYIMENESDKYFYFNPETKACK 2233  
 Qy 613 -----DKELY-KEDII--- 622  
 Db 2234 GINLIDDIKYFDEKIGIMRTGLISPNNNYFNENGEMQGYINTEDKMFFYFGEDGVMQI 2293  
 Qy 623 -----VPAGS-----TSNGPRIDL---LLKPD---VSAPGNKIKST 652  
 Db 2294 GVFNTPDGPKYFAHQNTLDENFEGBSINYTGW---LDLDEKRYFYFTDEYIAATGSVI--- 2347  
 Qy 653 LNVINGKSTY 662  
 Db 2348 ----IDGEYI 2354

ID AAB51367 standard; Protein; 3132 AA.  
 AC AAB51367;  
 DT 10-APR-2001 (first entry)  
 XX Japanese yam mosaic virus JYMV-M protein sequence SEQ ID NO:1.  
 XX Japanese yam mosaic virus; JYMV; Dioscorea; resistance; JYMV-M;  
 KW Dioscoreaceae mosaic virus.  
 XX Japanese yam mosaic virus.  
 PN JP2000300270-A.  
 XX 31-OCT-2000.  
 XX 21-APR-1999; 99JP-0113671.  
 XX 21-APR-1999; 99JP-0113671.  
 PA (AICH-) AICHI KEN PREFECTURE.  
 XX WPI; 2001-141353/15.  
 DR N-PSDB; AAF26998.  
 XX Weakly toxic strain of Dioscoreaceae mosaic virus used in the genetic  
 PT diagnosis of the mosaic virus infection and for the management of crops  
 PT resistant to the virus -  
 XX Claim 1; Page 6-20; 22pp; Japanese.  
 CC The present sequence represents the Japanese yam mosaic virus (JYMV)  
 CC JYMV-M protein sequence. JYMV is a weakly toxic strain of Dioscoreaceae  
 CC mosaic virus. The present invention also describes: (1) a weakly toxic  
 CC strain of Dioscoreaceae mosaic virus having a nucleic acid sequence  
 CC which shows interference against the infection of a strongly toxic  
 CC strain of Dioscoreaceae mosaic virus; (2) a Dioscoreaceae mosaic  
 CC virus-resistant crop which is a plant of Dioscorea japonica, Dioscorea  
 CC opposita or Dioscorea alata and is infected artificially by a weakly  
 CC toxic strain of Dioscoreaceae mosaic virus; (3) a method for the  
 CC cultivation of a Dioscoreaceae mosaic virus-resistant crop in which the  
 CC weakly toxic strain of Dioscoreaceae mosaic virus is infected to the  
 CC tubercle or the bulb of a plant of Dioscorea japonica, Dioscorea  
 CC opposita or Dioscorea alata and the tubercle or bulb is grown; and  
 CC which a DNA fragment amplified by using reverse transcriptase-polymerase  
 CC chain reaction (RT-PCR) primers atagttacacacccctcac (AAF26999) and  
 CC caccdtacattgcwgaragyg (AAF32201) is treated with a restriction enzyme  
 CC to distinguish the weakly toxic strain of Dioscoreaceae mosaic virus  
 CC from the strongly toxic strain of Dioscoreaceae mosaic virus.  
 XX Sequence 3132 AA;  
 SQ  
 Query Match 27.0%; Score 1186.4; DB 22; Length 3132;  
 Best Local Similarity 9.9%; Pred No.1.3e-08;  
 Matches 302; Conservative 170; Mismatches 180; Indels 2386; Gaps 132;  
 Qy 1 YP-----VVLADTSSSEDALNISDKV-----AB-NK----- 27  
 Db 26 LPPVFIGIPATSKTIATDKNDHEVHVIQSKSTEAQKKMNVNNAKFIHKHAEVYKQCAVL 85  
 Qy 28 --EKHEN-----IHS-AMETSQDFKEKKTAVTKEKVV-----S 58  
 Db 86 KWEIEYNTTRSTNPSSYKVLVTQOKSAVHRYNMEKARK-KLKEQEDILNKCIIGPWGPS 144  
 Qy 59 KN-----PV-----IDNNTS----- 68  
 Db 145 MNFASIASGPLPSAMEAEELKWLPHQTSQRISQIKGTPKQVTLGRGEFAKFKVNLTLMTQK 204  
 Qy 69 -----NEEAKIK-----EENSNSQSGDYTDSF-----VNKNTE 96  
 Db 205 SLLLELCGKHVRVCRREHKVKYLKINTKHEEGFNKARDVVMDFNTQRLLELMITRTSG 264

QY 97 N-----P-----KXEDKV-----YI 107  
 DB 265 NNRHSVONIKRPHSGVLNRETLGCTQSRAYGRVPIVRGNHSGKLYDARIKLSQIRRKI 324  
 QY 108 AEFKD-----KESGE----- 117  
 DB 325 VRFADPGVFMNGFNTAFQRYRKODREHTCETDLDBEEGCEVAALLCLALFPCKITCNK 384  
 QY 118 -----KALK 121  
 DB 385 CVEENLLSEGOATHENILKKOEIRHIVLQRHPOFKHALQILERSKALQSVNSNYKQFT 444  
 QY 122 ELSLKNKTVLYTYDR-----IFNGSAI-----ETT----- 147  
 DB 445 EIHSLSECKTLPAFQOARINDVILKGGSATBELSEATRNLEIIVRLKARTSESSEKGT 504  
 QY 148 -----P-----DN-LDK-----IKQIESISVE 164  
 DB 505 LKTFNKTSQKALNPALMCDNQDENGNFVWGERGYHAKRFENKYFELIDPSKGYAKPE 564  
 QY 165 -----RAQ----- 167  
 DB 565 ARINRGORKTAITRLIVPTNEVLRQEQGESIGENPLTVBCTSVLNGDPLFPCCCVTN 624  
 QY 168 -----KVQ-PMKNH----- 177  
 DB 625 EAGEPISELQMPKQHLVNGSGSKYVDMPQEGQMYIAKAGFCYMNITFLANLVNR 684  
 QY 178 KE-----IGVEE----- 184  
 DB 685 KEBAKAFKQWADVILNQGTPWPLLDVASACYLLKVFPPDSASLEPRIMVDHKTMTM 744  
 QY 185 AIDYKXINA-----PF-----GKNPD-----GRGVISN-IDTGDYR-- 217  
 DB 745 VYDSGSLNTGYHILKANTVEQLIFTRAGLSDMKHYLVG-GPILNEDIDP-TEYRTP 802  
 QY 218 -H-----KAMRID-----DDA 227  
 DB 803 SMLRLRIKGIYNPQVLLDDIRIDRIYLPYALLSPGVLIAMYSASLEILTRERYLRKDE 862  
 QY 228 KAS----- 230  
 DB 863 FVSIVLILESARKVSVSTLSQMLIEGEAQYIIIAVQGIKQRYPIPTYVMEMLIL 922  
 QY 231 -----MREKEDLKGTDKY-----WLSQKI PHAF----- 255  
 DB 923 ASRSESDALDAAGFKKFORBSIOLMEKUYLRILEDEWRRELSLRORFSATILRSSKFAMRT 982  
 QY 256 -----NY-----NG-----GKIT----- 264  
 DB 983 HGGELNNAISIEDIGRYSESMMNYFGEELKQGVNKIYKLTNOAKVITOSTHTSIKRYVSC 1042  
 QY 265 -----VEKYD-----GRDY----- 274  
 DB 1043 FNYLIPDVSKPINWVCLMILTLMQELHTWERTNRCKRIARRENEQEKHKIKPMQO 1102  
 QY 275 F-----DP-----HGMH-----IAGILAG 288  
 DB 1103 FONEHKVDPTEFEFLEYLCKHTPELLTYFOEDEVVVHQAQRGSELEBRVAFIALVMV 1162  
 QY 289 NDTQD-D--IKNFN-----GI-----DGJ--APN 307  
 DB 1163 FDSERSDCVVKILNKLKNIISTSDADVYHGLSEIEDIENKMLTIDELASODGVTRDPN 1222  
 QY 308 A-----QI-----FSYK----- 314  
 DB 1223 VMEHTFSTMWMTQVNSNGRTIPIHYRTGCHFMFTTRANAHVATEIATNEHKDMLMGAVGS 1282  
 QY 315 -----MYS DAG----- 320  
 DB 1283 GKSTGLPHLSKRGKVLVEPTRPLAENVYRQLSHEPFIYATLIMRGLTTCGSSPVTIM 1342

QY 321 -SGFA-----GD 326  
 DB 1343 TSGFALNQLANRRRIAEYDVFIDECHVDANAMALRCLLHDAEFPOKVIKVSATPPGR 1402  
 QY 327 ETMF--HAIE-----DSIKH-----N-VDVVS----- 345  
 DB 1403 EVEFTQHPVLLTEETLGLKEFYDAQGTGVNCVIRHGDNLLVYVASCNEVDIISKALI 1462  
 QY 346 -----VSSG-----FT-----GTGLV-- 356  
 DB 1463 DKGHVTVKVDGRTKYKGVKVEIITSGTPQRKHVVAATNIENGVTLDIEVVVDFGTKVVPF 1522  
 QY 357 -----GEKYQOAIRALR--KAG-----IPMVA----- 377  
 DB 1523 LDVONRMQYOKVAIINGERIORLGRVGRHKAQTALRIHTBERGLSEVPSCIALTEAARFC 1582  
 QY 378 -----TGNVATS----- 384  
 DB 1583 FTFGLPVITNNVTLSLSNATVROARTMAHPELSPFYTHFVRVYDGTWHPEIHKVLKRFK 1642  
 QY 385 -----ASSSS-----WDL 392  
 DB 1643 LRDEIYLNKTAIPNRGVNTWTSAYORLGANVDSNEIRIPLCKEVPETLHETIWDI 1702  
 QY 393 VANNHLKMTDTG----- 404  
 DB 1703 I-TTH--KSDAGFGRLSASACKVAATYLTQDVMISIORTIHIDALIYEBROKOEYFRIT 1759  
 QY 405 -----NVTNR--TAAH--EDAIVASAKQTVBPKVNI----- 433  
 DB 1760 TNSISSNFSLSQSIANAIARFSSDHTVENISVENAKAQCEFRKNLNIIDAAPODPOSOV 1819  
 QY 434 -----G-----GESFKR-----NIGA----- 445  
 DB 1820 GRSYISNEGALDAYHOSSEKAMSEHFKLKGMRNPLITRDMIMIGVLLGMMVYKQFK 1879  
 QY 446 -----FPD--KSKI--TTNEBGTK----- 460  
 DB 1880 SQMTEVHNHAKGKQROKORLRFRRARBSKLGREYVGGDGTMEHYFGEAYTKKGTSGKXH 1939  
 QY 461 -----AP--SKLKFV----- 468  
 DB 1940 GMGQKQRFVNMYSFDPEDFAVRFDVLTGATLDEPTITDHLVQEHFTKIRSEMIBSG 1999  
 QY 469 -----YTGKG----- 473  
 DB 2000 ELBSQHLVSGGVAVAYNNNRGTGALQVDTLRPNPLVLCANKPTIAGPEREYELRQTQ 2059  
 QY 474 -----OD-----OD-LIGLDR 484  
 DB 2060 PKAISLDVPRKANDLSEKVOHESASLHRLGRDVPNPISNNICKLINRSEGERDVTWGLGF 2118  
 QY 485 GKIAVMDR-----TY-----TKDLK----- 499  
 DB 2119 GPVITTRNHLFENHNGEGLDITRHHGDFLTMTWTKLOLYPVNRRDLILRLPKDIPRPPQK 2178  
 QY 500 -----NAFKKA-----MDK-----G 509  
 DB 2179 LQFROPERNEKICMVGNSFOAKSVNTVSETSILPMDDCHFWMGHIITTKDQGGCLPLVS 2238  
 QY 510 ARAIMVY-----NTVNY-----YNRDN----- 526  
 DB 2239 TRDGNIVGHSLSGFPNNITINFASF PENFVNQYLLTPBNHOWIOMWKYNTDNISMGALKI 2298  
 QY 527 -----W--TEL----- 530  
 DB 2299 SNEAPGLFKTKLIGDLESIFVROQMKREKMYVTOLDGNKAIASCPNOLVSKGVVKGK 2358  
 QY 531 -----PAMG----- 534  
 DB 2359 CPMFDWYLKLDAGRKFTPLGQYOKSRUNKKAYIIRIDIMKYSTVIEAGVQPOTFEDAV 2418  
 QY 535 -----YEADE-----GTSQGVF-----SI--- 548

Db 2419 RLLIQLSELGFCQVITDEDFVFNALNMKSAVAGLYGGKKDYFKDFTQEMKETILKQ 2478  
Qy 549 -----SGDDGVGLWN-----MINPDKTEVVRNNK-----EDPK- 577  
Db 2479 SCARLYTGKG--LWNGSLKAEALRPLEKV---QANKTRTFTAAPLDITLLGGKACVDDFFN 2533  
Qy 578 ----- 577  
Db 2534 QFYELNKGPNVGMKPYGWNELLTKLPDGIHCDADGSDQFSSLSPLYINAVNLRL 2593  
Qy 578 -----DKLEQ-----YYP-----D-----MESF-----NSNKP-----NV----- 598  
Db 2594 HFMETWIGQMLNLYTEIYVTPIDPDGTIVKKFGNNSGQSTVVVDNLTMLVLLALKY 2653  
Qy 599 -----GPEKE-----ID-----PK-----FAPDT-DK 614  
Db 2654 SLLKDGVEAEKHKQVIKYFVNGDILLISIDPAYEGLLDTMQGNFKELGLKYDFNSRTRDK 2713  
Qy 615 -ELYK-----EDIIVP-----AGSTSWG----- 631  
Db 2714 GELFMHQGRVEDIWIPLQEIRIVSILEWDRSKPEGRNMEALCAAMIESWGHELT 2773  
Qy 632 -----PRI-----DLILK 639  
Db 2774 QIRRYAWLQOAPYSGLAETKAPYIAESALRKLKLDKADQSAIEVYLRAIFEDYITE 2833  
Qy 640 P-----D-VSAP-----GK----- 647  
Db 2834 PEDLVVYHSGDDTLDAQTSAPSKARKQESASSTQGIAPTMEGPEPADDPGKSKQQTVI 2893  
Qy 648 NIKSTLV----- 655  
Db 2894 NLEKDVNVGTGTFAVPRLKGLATKMSMPRVRGKAAMNLDHLLVYNPQVDLANTRATRK 2953  
Qy 656 -----ING----- 658  
Db 2954 QFDTWYGVKRDYELDDSSMQIILNGLMWCIENGTSNPNMGWVMMDGEEQIEVPKPL 3013  
Qy 659 ----KST-----YGY 664  
Db 3014 IDHAKPTFRQIMAHFSYVAEYIEKRQEKAYMPRYGL 3051

RESULT 9  
ID ABG23329  
XX ABG23329 standard; Protein; 6619 AA.  
AC ABG23329;  
XX 18-FEB-2002 (first entry)  
XX Novel human diagnostic protein #23320.  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX Homo sapiens.  
XX OS-  
XX WO200175067-A2.  
XX PN  
XX PD 11-OCT-2001.  
XX 30-MAR-2001; 2001WO-US08631.  
XX PR  
XX 31-MAR-2000; 2000US-0540217.  
XX PR 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HYSEQ INC.  
XX Drmanac RT, Liu C, Tang YT;  
XX WPI; 2001-639362/73.

DR N-PSDB; AAS87516.  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX Claim 20; SEQ ID No 53688; 103pp; English.  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (II) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG0010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 6619 AA;

Query Match 26.7%; Score 1175.3; DB 22; Length 6619;  
Best Local Similarity 9.4%; Pred. No. 3.8e-07;  
Matches 345; Conservative 151; Mismatches 142; Indels 3047; Gaps 153;  
Qy 2 PV-----VLAD-----TSSSEDALNISDK-----E 21  
Db 2746 PVKATDVKPIKTKWEILKDLWNQFLKRAETNEAITLADEIIAAKGCKKSWASLFGSGD 2805  
Qy 22 KVAE-----NKEK----- 29  
Db 2806 MVQKLIDARANKPKPEKAVKODKTVPKKTVTTRSNADVIRSVLNDKMGSDYCKAHAIH 2865  
Qy 30 -----HENIHAME----- 38  
Db 2866 DVKRAAKSYAALGLTPCECSQIDKLKHLKIHTMAKYVRANKRAHDDDDIIRMNVAVLGL 2925  
Qy 39 -----TSQDFKE----- 45  
Db 2926 GTIAKTGLGVHPTVTTLRLSLGIEPADTRRTFMENVLAPLPTHVADWLSEQVGPKYEIRQ 2985  
Qy 46 -----KKTA-----VIK----- 52  
Db 2986 YVRDLLLLEYNNRHNHKETAHERFLKRYVEVVEMLLEELAPNSEBAYDTLHNAIAALTELS 3045  
Qy 53 -----EKEVVSKN-----P-----VIDNNTS----- 68  
Db 3046 ELMKSSDKVEYAPNSNRLAILDALADQIVTATGIGTFLGMNVPGALTEVNRNSYKSFEDG 3105  
Qy 69 -----NEEAKIKEE----- 77  
Db 3106 EPVFNENKVMKEAAAEAGFFDLDLPNASELRISGPGTGKTFTHMAHIDEIMPRYHETC 3165  
Qy 78 -----NSNKS-----QG-----DYTDS----- 89  
Db 3166 SLMGIPALYNEVIMTATTNKAEEVLARATGRPTSTYHSFQGLTVRNNLTKTGADLIIPSKS 3225  
Qy 90 -----FVNKNTENPKKEDKVY----- 106  
Db 3226 FSIKKNKVIIFIDEASMDIRLKLKFFINEGTH-----QSKIYVLGDKCOLLPVKETSSPVYND 3281

QY 107 --IAEF-----KXSEGEKAIK 121  
 Db 3282 SSIKEFMLTQOWRTDVPBELHALHEQURGTIEGKHGFLPIKAVPGIIMDKD---GDBMRK 3338  
 QY 122 ELSSLKATK---VLYT-----YD--RIFNG-----SAIET--TPDND 152  
 Db 3339 EVEGHFMKTDTRIYVAITNOQVNVNFTFRANNGHGEFCIOGEBLVNSAIOIGADRIL- 3397  
 QY 153 KIKO-----IEGI-----160  
 Db 3398 SIEQECQIVDDOSATRKILIDSSGIELEVRDCTLDTGCGVEGIPIPVMDYFIRLOKY 3457  
 QY 161 -----SSVERAOK-----VCP-----171  
 Db 3458 YAHQKWBHRYLKEFPPELRALHASTVHKSGSSTYDTIFIDATDLSGCRQPDVAVRLY 3517  
 QY 172 ---MMHAR-----KEIGVEBET-----D-YLK 190  
 Db 3518 VAVSRARH-RVVFYSLAEKFGELMOLG-EKOIREMTNSELVQELFLAELRLDKYLD 3575  
 QY 191 SINAPFGKNF-----DGRGVISN--ID-----211  
 Db 3576 SITNGETFAVYNSTGFLYLCEFYSSRSQOPSGHRLALANHLDDMREYLEQAGL 3635  
 QY 212 ---TGTDYR-----H-KAMR-----IDDDAKA 229  
 Db 3636 LLEVHVNQWVRLVOGCKTYDVARDALPECLIANDKOCHLSLQRTKAAFTLNDMA 3695  
 QY 230 ---SMR-----FKKE-----DLKGTD--KNY--WLS 248  
 Db 3696 LRHMRSGRLSNTMBEPLSEFCVCMRYITFNNEGIYPIALITTLTKDKDIVAVAMEPMDV 3755  
 QY 249 DK-----IPHAF-----NY-----258  
 Db 3756 DDOSMILELHATPAKPKKTPAKMOEFTTQELVP-AFEAOTQOYIICGSEYVQKLTQOG 3814  
 QY 259 ---NGC-----KIT-----VEK--267  
 Db 3815 KADANIGVMDCAVGNOKIYVPNYRQIFYPDPDKITKIALSMOALIDVSGSYVEPGTK 3874  
 QY 268 ---YDDG-----271  
 Db 3875 IHHVADYLPRAISKMLDLEMDVPLAVDIEFSLKHVDCGIGITFCMNHGIAFP 3934  
 QY 272 ---RDF-----275  
 Db 3935 VDYEPIEGATEAPYGRQVHNMLVRSMLRDFIKYLRQWYNHIAFDYVALIYOLEMTDL 3994  
 QY 276 DP---HGMI-----AGI-----LAGN-----289  
 Db 3995 DTEGLHGMISMLRWMDCTKLITYLATNSCAQNKLSLKDQAOEYAGNYAOEINDITRP 4054  
 QY 290 ---DTE-----292  
 Db 4055 LAELLEYNLDGLCTWVYVEKMDTLVNDQQLDVYTNIFKPCACEDIIQWLTGMPINRT 4114  
 QY 293 ---ODIKN-----FN--299  
 Db 4115 VLEVEALTDYNNALKTIAOSKVIKDFTRLNBEWEKONQILKKKVTLADAKEQENP 4174  
 QY 300 ---GI-----301  
 Db 4175 NSGIQOLKLEFFLGLVGLTASKLPRATSGSILSKMHTODTJILBELALIDYKAVD 4234  
 QY 302 ---DG-----IAPNAQ-----309  
 Db 4235 KILTAIPALKNAQOPGCMHYLFGNLNGTIVSGRLSSSEPNQLNPSGSRYAKMIKC 4294  
 QY 310 ---F-----SYKMSD-----318  
 Db 4295 FEAPGCMIFCGUDFASLEDRIISALTTKOPNKLVYTDGVDGSHLRAVAVFGEMMDIEDT 4354  
 QY 319 ---AGS-----GF-----323

Db 4355 VESVNSIQEKYKAYRODSKAPTFALTYQGTIVITLMKNGCFPEQKARHVEERYHTLYKUSD 4414  
 QY 324 -----323  
 Db 4415 DMVQAKLDQAAKDGVTVAFLRVTRPQLQOIVRGTSKTPYEABEGRTAGNALGOSWCL 4474  
 QY 324 ---AGDE-----TFPHA-----IED-----335  
 Db 4475 LNNRAGSEFMRKVRNSKRLDIRPSTHIDAOYFLIRDDMDVYITTNHLYKANQWODHP 4534  
 QY 336 SIKH-----339  
 Db 4535 DIAHPDVHLDDQEGAVIVQNTMLLTBEKQVTFRDLGRAQSALFQNLNDRFGQVVDVRD 4594  
 QY 340 ---NV-----DVV-----344  
 Db 4595 FLNVSYLGHMSQPDFOQNLVTNPQRMOPRYOABCECDIIQALNMTFDEGCLFKALMRNA 4654  
 QY 345 -----SVSGCF 350  
 Db 4655 ARLNGKPGNNAVYDAEKMTHYARIFIKESGGOKKEABEPVSALKKGBEIHMSSELF 4714  
 QY 351 T-----351  
 Db 4715 TLLTGEABEGVMSLAGHHTVPRRLGVMPIAVAMSKAEQVDVYRADGTTITHVPRHEID 4774  
 QY 352 -----351  
 Db 4775 WSDKIAPRATIANRFTWMLHDEYDVYKDKYLSVTLPLKIKHIVMKHRVDSVESMD 4834  
 QY 352 ---GNGL---VEGKYQOAIR---ALRKAG-----IPM 374  
 Db 4835 IMDVSTSMGTGLHDSLEKAWNHGHNALRLKLGPERVVQVNVINPTPEQIASPNLIPV 4894  
 QY 375 ---VWATG-----379  
 Db 4895 WIEQATKIKGWTIGCFDITVEGLLDQVSTSTYTMVKGGRDEHRLQOSMYRWLHDD 4954  
 QY 380 ---NY-----ATSA-----SSSW-390  
 Db 4955 KITEVIRINYIFTDMMKALATSOENYPKHVLHKDIPLMSYATEEWIKAKLALIDKYW 5014  
 QY 391 -----390  
 Db 5015 DAPSEIPECTDEBLMSDPVFKYPADEKAKOPGASTKNPPTLFEARKFMAEKGTI 5074  
 QY 391 ---DLVANNHLKM-----TDTG-----404  
 Db 5075 LHVEPIFPMIDLTVGSHHRALEIYDVULCNKTONTDRGFPVEVAYFLAKMA SCMGATII 5134  
 QY 405 ---NVTRTA-----AH-----EDAI-AVA-----419  
 Db 5135 TKDRGEIPVNIYAVALATSGFGKHSVNIVEDGFMAGFKRPMEDTMRALBRLMKIAN 5194  
 QY 420 ---SAKN---QTVEDPKVNIAGESFKYRNIGAF--FD-----448  
 Db 5195 ERSARNATDOKEBDKV---EA-EYRTGAYPFTFOSGTRPRAVKOLYHNKLLAGCGAIN 5249  
 QY 449 ---KSKITTN-----E-DG-TKA-----P 462  
 Db 5250 LOIDEIGSNLLANADVLTFLVLELYDQGIKIKLTKRTAESTIRSEEVDSKTFANLLPSTP 5309  
 QY 463 SKL---KFVY-----IGKG-----OD-----475  
 Db 5310 SKLNGSQTEELFYFDLTGYSRRCLPAVGQADKRAVLSQABEYRNLIKQDNNAVYVK 5369  
 QY 476 ---ODLIG-----480  
 Db 5370 WANHFHSLADSKFGKMWKTVEDDVGIALIDYKIQCEKQOAAALADHEIRKAELEHRYFKA 5429  
 QY 481 LDLRGKIAVMD-----RIYTK-----496

Db 5430 LKAGAFAPVDSINVMEHLKQAILLVEESGVAFTQILNREKAYVVKLAKYIASVGTVEV 5489  
Qy 497 --DL-----KVA-----FK-----KAMD----- 507  
Db 5490 HADLLEALPFYKSGNAARNEMMTLATAMGYKQHIIKKTYNGBIEFFRGELKETDTNEM 5549  
Qy 508 -----KARA-----KGAR-----IMVVTWN 520  
Db 5550 IVSYGENFAYDIABKVPDQLHVLVTOQAGYHAWNHFRGRHAEENALAGNFMIVIDVD 5609  
Qy 521 -----Y-----NR-----DN-----WT 528  
Db 5610 GTCSLDVCHLMKEYRFTYTKRHTDEENRPLIIPMNYQLELDGAEYKEPFMDNMWAW- 5668  
Qy 529 ELPANGYRADE-----GT-----KSQVF-----SIS 549  
Db 5669 -LP-----FKTDESANQRAKWKESCAKGYTYNNMDAPLLDVRDFIPRTSKNEAFQOQMKVQ 5724  
Qy 550 GDDGVKLV-----N-MI-----NP-----DK----- 564  
Db 5725 NLONLIERPFAQRIATGNRNHMIKVALALVDSGIGFNEVKQVRHEFNKLSNPLSADEL 5784  
Qy 565 ----- 564  
Db 5785 STVMITVAKLLFLVBEDCOLKEETHSEBEHGHGEMNDQLILIVGSAAGKSASLRNROE 5844  
Qy 565 -----KTEV-KR-----NNK-----BDF-----KDKLE 581  
Db 5845 RNVYLNTAGKRLPFKNFNGVRITDPYQILEYFDQCIEHRDDVDGIIVDSLTFMWMLE 5904  
Qy 582 -QYY-----PI-----DMESEF----- 591  
Db 5905 TOYVLTANTQKAMGEFAOFFKILLQEKVKVFAKPVIFTAHVKDEVDERAMELKTVPVK 5964  
Qy 592 -----N-----SN-----KP----- 596  
Db 5965 GSKKNGIBAYFSTVSAERIDLKELEKYSNGMLEITDEQELGYKHVFQTRPTKTKVCK 6024  
Qy 597 -----NVG----- 599  
Db 6025 RIRSPMGFSKOETYMNDNAQKLLDHLAEYYSTGOAAKAEKATDNLGGFGAKESDIYLA 6084  
Qy 600 ----- 599  
Db 6085 NLKAYAGKAASGANFIQIIADLTDLHGSAGEYREQLYITSGTEKGCKCTYEKNKGEYF 6144  
Qy 600 -----D-----EKEI-----DF-----KFA----- 609  
Db 6145 LPGYTVINDILVMTSGETIPEAVFEKVVNVYDFDEKKEVAKSVMPVNAIGKFAVAIL 6204  
Qy 610 -----PDT-----DKELYKEDIIV- 623  
Db 6205 KSEEDKQTKDGSNGYVSTGETRFTNTIEKVFHPDLHLTVVEABELTERGKELTVEEAVFM 6264  
Qy 624 -----P-----AGSTS----- 629  
Db 6265 DKWLEKNKGVTRDKTKTGKASGAKGQPPKPGATNTGAGASAASLFEIMQIIVLGVDSRF 6324  
Qy 630 --WG-PR-----IDLL-----KPDVSAPG-----KNIKS-----TLNVI----- 656  
Db 6325 RNWGLARGMLDLETGLSLGLDLKLVTETPD-----GTKQVRQNSKOMQAEDITTGVIDWF 6380  
Qy 657 -----NKSTVGY 664  
Db 6381 KEAKVIFVEVPVQSQANGMKSYGV 6405

RESULT 10  
ABG23295  
ID ABG23295 standard; Protein; 5909 AA.  
XX  
AC ABG23295;  
XX

DT 18-FEB-2002 (first entry)  
XX Novel human diagnostic protein #23286.  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX Homo sapiens.  
XX WO200175067-A2.  
XX 11-OCT-2001.  
XX 30-MAR-2001; 2001WO-US08631.  
XX 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HYSEQ INC.  
XX Drmanac RT, Liu C, Tang YT;  
PI  
XX WPI; 2001-639362/73.  
XX N-PSDB; AAS87482.  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX Claim 20; SEQ ID No 53654; 103pp; English.  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX Sequence 5909 AA;

Query Match 26.7%; Score 1173; DB 22; Length 5909;  
Best Local Similarity 9.2%; Pred. No. 2.7e-07;  
Matches 311; Conservative 169; Mismatches 153; Indels 2730; Gaps 130;  
Qy 1 YPVVLADTS-----SSDALN-----ISD-----KE----- 21  
Db 2305 YAVWITDVPDRNHRIVHVIRESGRTSEYVNNLLAQTPEKVPDLTETQAKFVRL 2364  
Qy 22 -----KVAE--NKEK-----H----- 30  
Db 2365 NNTGVISVYNKLEDGIPVKVAVKATEPTNKEKLEAASTKLKMGAVATDAYEFTDLNE 2424  
Qy 31 -----ENIH-----SA-----METSQ-----P 43  
Db 2425 NEETIVAVNDKVPDQVNLHRRHRSANLKDYGFTKFLERLSTVIOKRRHSVEDLMKF 2484  
Qy 44 KE----- 45

Db 2485 MEKGDLPADGDSIVI FKRLKESKDNPEHLKQVMDDDVFDCHSKKIEQCVRMYKENVL 2544  
 QY 46 ----- 45  
 Db 2545 DQNRKDCSHGLHVASLOYIRNFGNVTITIGVABEDVFAVEYDVTQRVSAHIITKL 2604  
 QY 46 ----- 45  
 Db 2605 PDALRDHVNNGNPISTIEGCTELMMVLSGNHPERSQOVUGGHYTNLTYSELTRNYLD 2664  
 QY 46 ----- 65  
 Db 2665 DDEDSVIPDEPVIKKKTALNMEESFATNPKAEPVATDVKPIKTKMELKQLMNQFLKA 2724  
 QY 66 NTSNE-----EAKIKEENS-----NK----- 81  
 Db 2725 ETSNEAITLADBEIIAAKGCKCKSMASLGFGDMQOKLIDARANKKPEKAVKODXTVAPK 2784  
 QY 82 ----- 85  
 Db 2785 KTTVRSRNVADVRSYLDKQMSDYCKAAIHIDVGRAAKSYAALGLTPECSQIDKLKH 2844  
 QY 86 ----- 95  
 Db 2845 HLKHTMAKVYRANKRAHDDIIIRMNAGLSLGTIAKLTGVHPTVTTLRLSLGIEPADT 2904  
 QY 96 -----EN-----PKKE----- 101  
 Db 2905 RRTFEMENVALPLPTHVADWLSEQVGPKEIRQYVBDLLBAVNNBNHKEHAFELKRY 2964  
 QY 102 ----- 107  
 Db 2965 VEUVMELEBLANSEAYDTLHNAIALTELSELKSSDKETAAVNSNRNALILDALDOI 3024  
 QY 108 ----- 124  
 Db 3025 VTATGTFGLMNVPGALTEVNRSNYSKDEGEPAFENENKMYKMAABEGFFDLDPNA 3084  
 QY 125 ----- 133  
 Db 3085 SELRISGPGTGCTFTMAHMIDEIMPRYHETCSLGIIPALYNEVIMTATTNKAEBVLARA 3144  
 QY 134 ----- 137  
 Db 3145 TGRPTSTHSGOGLTVRNNLKTGBADLPSSKFSITKQKVIFIDEASMDRKLKFFINEG 3204  
 QY 138 ----- 143  
 Db 3205 THQSKIYVLGDKCQQLPVKETSSPVYNDSSIKERFMLTQOMRTDPELHAEQLAGTIEG 3264  
 QY 144 ----- 174  
 Db 3265 KHGFLRPIKAVPGIIMIKDQDEMRKEVEGHFMTKTDTRIYATNOQVNVNTPFIRANMG 3324  
 QY 175 HARKSIGVEE-----AIDYLK-----SIN 193  
 Db 3325 HGEPTIG-BELVNSAIOIGACDRLSIEOEQIVDQDSATRKILIDSSGIELEVRDCTLD 3383  
 QY 194 AFGKMPDG-----RGMVISNI-----DT-----GTDY----- 216  
 Db 3384 TGYGVFEGIRIPKPEFPRLALHASTYHKSOGSTYDTIFIDATDTLSTCRQDVVARYLLY 3443  
 QY 217 -----RHK----- 222  
 Db 3444 VAVSRAHRVAVYGSIAEKFGGEFYSSRSQOPSGHRLALANHLDDMREYLEQAKL 3503  
 QY 223 ----- 229  
 Db 3504 LLEVHVNOMVRLVOGCKTYODVRDALPECLIAMDKOGLKSLORTKKAFTLLNDMA 3563  
 QY 230 -----SMR-----FK-----KEDLKGTQKMYLS 248

Db 3564 LRHNRKSCRLSNTWREPIILSFVGCARYITFNNEGIYPIALITTLARKDQIVKAVMEPMDV 3623  
 QY 249 DK-----IDHAF-----NY----- 258  
 Db 3624 DKDSVMIELHTAPGKKTPAKEMOEFTIOELVP-AFBAOTOYIICGDSERYKOLTKOG 3682  
 QY 259 -----NGC-----KIT-----VEK----- 267  
 Db 3683 KADANIGYVMDCAVGNOKVIYVPRYROIIFYDPDKIKTKIALSMQALIDYSGSYVEPOTK 3742  
 QY 268 ----- 271  
 Db 3743 IHHADYPLTPEAISKWLKLEMDVPLAVDIETFSLKHVDCGIGITIFCWNKHGIAFP 3802  
 QY 272 -----RDY----- 275  
 Db 3803 VDYEPIEGATEAPYGRQVHNMLVRSMLRDFEPIKYLBNOMYHIAFDVYALIYOLFMTDL 3862  
 QY 276 DP-----HGMHI-----AGI-----LAGN----- 289  
 Db 3863 DTBGLHGMISIMLANWDCSTKLITYLATNSCAGNLSLKDOAOBYAGNVAOBEINDITRIP 3922  
 QY 290 -----DTE----- 292  
 Db 3923 LAELLEYNLVDGLCTWYVEYKHMDTLVNDQOLDVYTNIPKPACEDIIOMQLTGMPINNRT 3982  
 QY 293 -----ODIK-----PNG 300  
 Db 3983 VLEVEEALTDYNNALKTIAOSKVIKQFTRLNEBWEKONQILKKKVTIADAKOQNP 4042  
 QY 301 IDGIA----- 305  
 Db 4043 NSGIOLKLEFELGLPVGLTASKLATPGSGILKSLKNTQDSILEILDALIDYKAVD 4102  
 QY 306 -----PNO----- 309  
 Db 4103 KILTAPIALKARQRPDMHYLFQNLNLGTVSGRLSSSEPNQLNPSGRYAKMIKC 4162  
 QY 310 -----IF-----SYKYSAGSGFADETMFHA----- 332  
 Db 4163 FEAPPGWIFCGLDPASIEDRISALTTPDKPNLKVYTD--GYDG-----HSLRAYAVFGE 4214  
 QY 333 -----IEDSIKHNVDSVSS-----GF--TGSG 354  
 Db 4215 MNPDIEDTVE-----SVNSIQEKYKAYRQDSKAPTALTYIQGYITILMKRQCFPEBOAR 4268  
 QY 355 LVGERKY-----W-----OAIKALKRAGIPMVV 376  
 Db 4269 MVEERYHTLVKUSDWQAKLQDAKQGYVTAVGLRVTRPLLOVIRGSTK--PYEAB 4326  
 QY 377 -----ATGNVATSASSSW----- 390  
 Db 4327 AEGRTAGN-----ALGOSWCLLNBPAGEFMRKVNNSKHRLDIRPSTHIDAQYFLIRDM 4382  
 QY 391 DLV--ANNHLKMT-----DTGNVTRTAHED--AIVASAK-----NOT 425  
 Db 4383 DVVIYTNHLVKAQVOMQDHPDIARPDVHLDDQEBQAVIYQNTMLLTKEKQVTFRDLGRAOS 4442  
 QY 426 VEPDKV-----IG----- 434  
 Db 4443 ALFQNLNDRFGQVVDVDRDIFVLNYSYLGHMSQRPQONLVTPQRMEOPRYOABECEDIIQ 4502  
 QY 435 -----GESFK--YRN-----IG-----AFDQSK-----ITTNEGC-----TKA 461  
 Db 4503 ALNMTFEGCULFALMRVAAARLNGKRGNNAAVYDAEPMVYANRIFIKESGQKGEABE 4562  
 QY 462 P-----SKLKFVYIG--KGODDOLIG----- 480  
 Db 4563 PVSAAKKGEIHNLSBELFTLLTGEAABGEVMSLAGHHTHVPRNLGMPPLAVAMSKAEQ 4622  
 QY 481 -----LDLRGKIA-----VMDRIY----- 494  
 Db 4623 VDVVRADGTTITVHPHEIDMSDKIAPARITAMRFTYWLHDEYDVYKDKKYSVTLL 4682



```
Qy 495 -----TKDLKNAFKKAMDKG-----ARAI- 513
Db 4683 KPIKIHVMKHRVDSLVSMDIMDLVSTSMGTGLHDSIEKAWHGHNAALRKLGYPERVVQ 4742
Qy 514 -MVNT-----
Db 4743 NVVINTPEQIASNPILPVWIBORATKGIAGWTIGGKFDIVTEGLLDQVKSSTYTIVWK 4802
Qy 519 -----VNYNRDNMT-----LPA 532
Db 4803 GGRDEHRLQSGMYRWLHDDKITEDVIRINIFTD-WMKALATSQENYPKRVHLKDIPL 4861
Qy 533 MGYEA-----DE-----539
Db 4862 MSYEATEBIKAKLALIDKYWDAPSEIPECTDELMRSDPVFKYFADPEKAKOPGARST 4921
Qy 540 -----GT-----KSO-----V 545
Db 4922 KNFTLFEARFKMAEKGKGTILHVPEIPFMDTLTGVSHHPAIEIEIVDLCKNTQNTDRGF 4981
Qy 546 FSI-----SG-----DDGVK-----555
Db 4982 FRVEVAYFLAKWASCMGATIITKORGEIPVNIYAVALATSGFGKHSVNIVEDGFMAGFK 5041
Qy 556 -----LNMNIPDKKTEVKRNN---KEDFKDLK-XY-----YPIDMESF 591
Db 5042 KRFMEDTMPALAEADRLMKIANNERS-----ARNATQKEEF-DKVEAEYRRTGAYPF---TF 5093
Qy 592 NSNKP-----N---VG-----DE-----601
Db 5094 DSGTPPAVKQRHKLKLLAGCGAINLQIDEISNLLANADVLTVFLELYDQSKIKQKLTKN 5153
Qy 602 -----KEID-----FKF-AP-----DT-----DK-- 614
Db 5154 TAESIRSEVDGKTPANLFLGTSPKLLNGSQTEELFYDFLDTGYSRRLFAVGQADKRA 5213
Qy 615 -----ELYK-----ED-----620
Db 5214 YLSOSAIEIYRNLIKQDNNVNVKWNHFLSHLADASKFGWKMTVEDDVGIALIDYKIQCE 5273
Qy 621 -----IIV-----623
Db 5274 QAAALADHEIRKAELEHRYFKALKAGAFVDSNIMWEMHLKQAILLVEESGVAFQ 5333
Qy 624 -----P---AGSTS-----W 630
Db 5334 TILNREKAYVKLAKYIASVGTEVTHADLLEALPFYKSGNAAQKVPFDQLHLVLTQAGYHW 5393
Qy 631 -----GPR-----ID-----LLKPD-----641
Db 5394 ANHFRGRHRAEANAIAGNMIVDVLFLVEEDCOLKEETMSEEFEGHGMNDLILIV 5453
Qy 642 -VSAPGK-----NI-----KSTLN-----654
Db 5454 GFSAAKGSASLRNIRNQRWVYLNTEACKRPLPFKNFNGVRITDPYQILEYFDQCIHRD 5513
Qy 655 -----VI-----656
Db 5514 DVDGIIVDSLTFMMDLQVVLTAANTQKAWGEBFAQFFKILLQEKVVKFAKPVIFTAHV 5573
Qy 657 -----NKGSTY-----662
Db 5574 KDEVDERAMELKTFVPVKSLKNGNIEAYFSTVVSAAERIDLKLEKYNGMLTEITDEQE 5633
Qy 663 -GY 664
Db 5634 LGY 5636
```

RESULT 11  
ABB58144

ID ABB58144 standard; Protein; 7107 AA.

```
XX ABB58144;
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 1224.
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX (PEKE ) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX N-PSDB; ABL02247.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX Disclosure; SEQ ID NO 1224; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABU16176-ABL30511), expressed DNA
XX sequences (ABU1840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 7107 AA;
XX Query Match 26.6%; Score 1167.8; DB 22; Length 7107;
XX Best Local Similarity 7.6%; Pred.No. 7e-07; Gaps 139;
XX Matches 324; Conservative 163; Mismatches 156; Indels 3642; Gaps 139;
Qy 1 Y-----PVVL---AD-----TSSSEDALNISDEKVA--ENKEKHENI-----33
Db 1071 YDVNVRGEPAPVTWYQNDKELKPEELPSSE-IGNIPYNTKISITVETVRKHTGIYKIIA 1129
Qy 34 ---HSAMETS-----QD-----PK 44
Db 1130 VNEHGQDEATVEVNIAPPSKPRGLDVKVTQKSKLKWKKPEDDGGKPI SAYQVKEKPD 1189
Qy 45 EKK-----TAVIKEKE-----VSKNP---61
Db 1190 KQGRWVPLGRTSANDTEFDVKGLOEGHEYOFRVKAINEGESDPLSDSDSIANKPYDA 1249
Qy 62 -----VID-----NNTSNEEA 72
Db 1250 ASKPGTPNIVDNEHMKLWEAPRSDGAPISGVIIIEKKDFSPIDWELSTNTSVPEA 1309
Qy 73 KIK---EEN-----SNKSGQ---DYTDS-----FVNK-----93
Db 1310 TVEGLVEGNIYQFRVAVNK-AGFSDPSDAPTEPHLAKPNLKPYNRDKMKPIKVRAGOP 1368
Qy 94 -----NT-----EN 97
```

Db 1369 VKFDVVKGBAPSLTWFLKETELSTGVRLENIDVNTKLTLDTPDRKOGQYKLRAN 1428  
 QY 98 ----- 97  
 Db 1429 INGVDAVVEV11DKPSRPEGLIEVSDIHKEGCKLWKRKPDGGLPIGTIVIEKMDTA 1488  
 QY 98 -----PKED-----KVY----- 105  
 Db 1489 TGKAVPAGSVDPKDIETIKGLDPNHRQFRVKA VNEGSESEPLETSATIAKAPFVISA 1548  
 QY 106 -----YIAEFKDESGE--KAIKELS-----SL 126  
 Db 1549 PRGLPELEDWDEHHVYKLMKEPPIRDGSGSPITNVIIEVMDKSGEFVKA VETDSPCKGV 1608  
 QY 127 KNTKVLTY-----DR1----- 138  
 Db 1609 KLEEGOOYKFRVAVNKA GSPDSEOTNMHVAKRPLKPHIDRVNLKPVITGLSISL 1668  
 QY 139 -----FNCSA1ETTP-----DNLDKIKQIEG1SSVERAKOY-----PM 172  
 Db 1669 DINRGEAPKVEWFFNNSVTSDEHSVKIDVINTKF-----FVMAKORSOGSKYIITKA 1724  
 QY 173 MNHARKEIGVEA-----IDYTK- 190  
 Db 1725 TN-----EYGEDEBALEVTVLGKPKRKPLOVNDITKHSCKLWKEKPPDDGSGPIDYVEI 1780  
 QY 191 -----SIN-----APF 196  
 Db 1781 EKLDPHTGOMLPCGKSTERPAKVIQLHEGKAVKFRVAVNKEGESEDELETERPIAKNPY 1840  
 QY 197 -----G-----KNF-----DG----- 202  
 Db 1841 DEPRRPGKRPPTNMDKDFVDLAMPKPDGAPLOQKVIQMRKDSGRAMVDSATVPDGC 1900  
 QY 203 ----- 202  
 Db 1901 NGVTGVEGHEYEFRIVAVNKA GSPDSVSKSVIAKRPFLKPHIDRKNLOKKIMRSGQ 1960  
 QY 203 -----RGMY----- 207  
 Db 1961 MLHIDALIKAPPAKVTWYNKTEIKTSDHIKIENEDYKTFIMPKVRADRGIIYVAK 2020  
 QY 208 -----SNI-----DTGTD----- 215  
 Db 2021 NDGSGDIVELEVLCKPSKPKGPLAVSNVTAETLHLKMEKPEDGDP1EQYLVEMDT 2080  
 QY 216 ----- 215  
 Db 2081 ETGRKVPVLTTKTPRADVTGLTEGKEYLFRVKA VNSEGESEPLEVTDTITKAKNPFDAADT 2140  
 QY 216 -----YR----- 217  
 Db 2141 PGKFOIVMSGNHCDLKWRAPEDDGASITGYIVERKDPNTGKMQALESTPDCARVN 2200  
 QY 218 ----- 217  
 Db 2201 DL1AGNKYQFRIMAVNKA GSKSPESDQMTAKOFAPPKIDRTN1KDIITKAGQHIFD 2260  
 QY 218 -----HKAMRI--DD-----DATA-----SMRF----- 233  
 Db 2261 IKVSGEPATKVMHLNKA RLENDSDSNVNI DMESYTKLTP1SKAFHSGKTYTLKAENSG 2320  
 QY 234 -----KKEEDL----- 238  
 Db 2321 RDEASFV1VLDKPPEGLRVTDVHKEGCKLKANAPLDDGLPIDHYIIEKMDVESGR 2380  
 QY 239 ----- 238  
 Db 2381 WLPGRFKESFAELNLEPSHEYKFRVLA VNTGESEPLETGEOSVIAKNPDEPGKPGTP 2440  
 QY 239 -----KGTDKNVL----- 247

Db 2441 EAVDMDKDHDVLRWPRPINDGSPITGYVVEKREKGTDK--WIKTEITITPCLGEBSKAT 2498  
 QY 248 -----SD-----KI----- 251  
 Db 2499 VPTLNCSEYEFVRKAINAAGPESDASKPIITKPKLAPKIDRKNIRTYNFKSGEPIF 2558  
 QY 252 -----P-----HAFNY--YN----- 259  
 Db 2559 LDINISGEPAADVTWNNQNNKSVOTTSFSHIENLEPYNTKYINNPERKDTGLYKISAHNFY 2618  
 QY 260 -----GK-----ITVEKYD-D-- 270  
 Db 2619 GQDVEFOINITKPKRPEGLEVESEVHKDCKLWKKPKDGDGEPVESYLVEKFDPTG 2678  
 QY 271 -----GR-----DY-F-----HGM 280  
 Db 2679 IMLPVGSDGPEYVNDGLVPGHDKFRVKA VNEGSESEPLETGS1IAKOPFSVPTKGV 2738  
 QY 281 H-----IAGIL 286  
 Db 2739 PEPTDTANKVELAMPPEPASDGSPIOGYIVEVKKYSPLEKALBETNSPTPTATVOGLI 2798  
 QY 287 AGNDTE----- 292  
 Db 2799 EGNFYQFRVALNKGSGSEPSDESKIFTAKPRYLAPKIDRRNLNITLSSGTLALDANI 2858  
 QY 293 -----QD1KNF-----GLDGI 304  
 Db 2859 TGEBAKPVEMKLSYHLQSGKNVTIETPDYTKLVIRPTORSGEVLYATNTSGKDSV 2918  
 QY 305 -----APNA--OI-----FSYKMSDAG----- 320  
 Db 2919 LVNVVITDKSPRPNGPLQISDVHKEGCHLWKRKSDGCGPIEFQIDKLEPETGCM1PS 2978  
 QY 321 -----SG-----FAGDETM--F----- 330  
 Db 2979 CRSTEPOVDVTGLSPGNEYKFRVAVNAEGESQVLVGDESI VARNPFDEPGKPENLKATD 3038  
 QY 331 -----H-----A1EDSI----- 337  
 Db 3039 WDKDHDVLA WTPPLIDGSGPISCIYIEKODKYGWERALDVADQCKAT1PDLVEGOTYK 3098  
 QY 338 ----- 337  
 Db 3099 FRVSAVNAAGTGESDSTPPIIAKARNKPP1IDRSSLVEVRIKAGOSTFPDCKVSGEPAP 3158  
 QY 338 -----KANTDVVSV-----SSG--F----- 350  
 Db 3159 QTKWLKKKEVYSKDNKVTNVDTNKLKVN SATRSDSG1YTVFAENANGESADAVKTV 3218  
 QY 351 ----- 352  
 Db 3219 IDKPA PNGPLKVDINSESCTLHMNPDDGCPIDNVVEKLEDTTGRW1PAGEIDGP 3278  
 QY 353 ----- 352  
 Db 3279 VTALKVGLTFGHHYKFRVAKNROGTSEPLTTAQA1IAKNPFPVPTKPGTPTTKDPKE 3338  
 QY 353 -----TGLVGEKY-----WQA-----I 364  
 Db 3339 FVDLEWTRPEADGSPITGYVEKRDKPSRPMKACBSIDITNAHVPLDLEGKYERFV 3398  
 QY 365 RALRKAG-----IPNV-----VATGNV-----ATSA--SSSS 389  
 Db 3399 RAVNKA GSPSDATETHVARPKPTPKIDNFMWDIKIKYAGNVFEPDVVPYTGAPLSKD 3458  
 QY 390 MDLVANNH-----LKMTDPS-----NVTRTAHEDAI 416  
 Db 3459 W-----THEGNM1INTDRVK1SNFDDTKIRILDAKSDIGSVYTLTARNINGTORHNVK 3513  
 QY 417 AVASA----- 421  
 Db 3514 TILDAPSVBGLRANGVSKSV1LWVRP PKDGDGSEITHVVEKMDNEAMRWVPVGDCT 3573

```
QY 422 -----KNOTVER-----DK----- 430
Db 3574 DTEIRADNLIIENHDYSFRVRAVNKOGOSOPITTSOPITAKDPYSHDPKQGPQATDWGKH 3633
QY 431 ----- 430
Db 3634 FVLEWSTPKRDXGAPISYIIIEKRPKGQWERAANVLGDCKAHVPELTNGGGEYFRVI 3693
QY 431 -VNIGGES-----FKYRNIGAFDKS----- 450
Db 3694 AVNRGGSDDPSSTIIICRFLAPFDKSLDLNDITVHAGKRLGWLPIEASPRPLITW 3753
QY 451 ----- 450
Db 3754 LYNGKEIGNSRSGSLFQNELTPEIVSSLSRSGRYTLILKNEHGSFDSAAHATVLD RP 3813
QY 451 -----KIT----- 453
Db 3814 SPKGPLDITKTRDGGCHLTWNVPDDGGSPILHYIEKMDLSRSTWSDAGMSTHIVHDV 3873
QY 454 -----TNE----- 456
Db 3874 TRLVHRKEYLFRVKAIVNAGESDPLEAVNTIIAKNEFDEPDAPGKPIITDWRDHDIDLOW 3933
QY 457 -----DG-----TKAPS-----KLKFFVIG--K 472
Db 3934 AVPKSGGAPISEYIIQKKEKGSPTWNVHRVPSNKNNTTIPELTEGOEYFRVIAVNOA 3993
QY 473 GQDQ-----DLI----- 479
Db 3994 GQSPSPSDIMAKPRYLPPKIITPLNEVRIKCLIFHTDIHFIGEPAPATWTLNSNP 4053
QY 480 -----GLDLRG--KIAVMDR----- 492
Db 4054 LLSNDRSTISIGHHSVVHTVNCORSDSGIYHLLLRNSSGID-EGSFELVLD RPGEPEG 4112
QY 493 -----TYTKDLKNAFKKAMDKGARIMV-----VNTVNY-----YN----- 523
Db 4113 PMEYEEITANSVTISWPKPKONGSGEISSYVIEKRLDTHGGGWPAVNVYSAKYNHAYVP 4172
QY 524 ----- 523
Db 4173 RLLEGTMYELVMAENLQGRSDPLTSQDPVAKSQYTVPGAPGKPELTSDKNHITIKWK 4232
QY 524 -----RD-N--WTEL-----PAMGYEAD-----EG 540
Db 4233 QPISNGGSPITGYDIERRDVNTGRWIKINGQVPVTAEYQDDRVTNSHQYQYRISAVNAAG 4292
QY 541 T-----KSQVPS-----ISG----- 550
Db 4293 NGKTSEPSAIFNARPLREKPRFYFDGLIGKRIKVRAGEPVNLNIPISGAPTPTIEMKRGD 4352
QY 551 -----DP----- 552
Db 4353 LKLEGRISYETNSERTLFRIDDSNRDSKTYVTAAENFGKOTADIEVIVWDKSPPE 4412
QY 553 -----GVKLW----- 557
Db 4413 GPLSYTETAPDHISLHWYSPKDDGSDITGYIIIEFTFEGVDWKPVPCPTCPNTNPTVKNL 4472
QY 558 ----- 557
Db 4473 VEGKYPVIRAEINYGASEALEGKPVLAKEPDPGAPSPQTTISAYTFNSANLEWHPPD 4532
QY 558 -----NM-----IN----- 561
Db 4533 DCGKPIGTIVERRERGGEWIKNNYPTNTSYTVSNLRDGCARYFRVLAVNAGPQHP 4592
QY 562 -----PD-----KTEVKRN-----NKE 574
Db 4593 SKPSDPMTAEHQRYRPDPPEPKPDRIITRNGVTLNWRPPTDGRKSGRIKGYVEMRPKNKG 4652
```

```
QY 575 DFK--D-----KLEBY-----YPIDME----- 589
Db 4653 DWKTVDIPINSTVYTVPSLSKEGEYSFRVVAENEVGRSDPSKPSQPTIIEQPNKPCME 4712
QY 590 -----SF-----NS 593
Db 4713 LGKVRDIVCRAGDDFSIHVPYLAPPKPNAFWYSNDNMLDDNNRVHKLHLDAAASVVVKNS 4772
QY 594 -----NKPNVGDE- 601
Db 4773 KRADSGOYRLQLKNTSGFDTATINVRVLD RSPPTRLRADEFSGDSLTYWNPEN--DDG 4830
QY 602 -----KEIDF----- 606
Db 4831 GSAIQNYIIIEKEARSTWSKVSFCTVPFVRIRNLVLNKEYDFRVAENKYGQSDPANT 4890
QY 607 -----KF-APDT-----DK-- 614
Db 4891 SEPTLARHPDIPNTPGIPHGIDSTEDSIITAWTKPKHDGSGSPITGYIIEKRLSLDDKWT 4950
QY 615 -----EL-----Y--KEDIIV----- 623
Db 4951 KAVHALCPDLCKIPNLIIENAEYFRVAANVNAAGSAYSGSSDLIFCRRPPHAPKITS DL 5010
QY 624 -----PAGSTSW-----GPRI----- 634
Db 5011 SIRDMTVIAGDEFPRITVPYHASPRPTASWSLNGLEVIPGERIKFDSNDYASMYNKS AKR 5070
QY 635 -----DULLKP-----DVS----- 643
Db 5071 DETGTYITLTNNKSGSDTASCHVTVD RPLPQGPGLNAYDITPDTCTLAWKTPLDDGGSP 5130
QY 644 ----- 643
Db 5131 ITNYVEKLDNSGWVKISSPVRNTHYDVMGLEPHYKYNFRVRAENQYGLSDPLDIIEPI 5190
QY 644 -----APGK----- 647
Db 5191 VAKHOFTVPDEPGPKVIDWDSGNVTLIWTRPLSDGSGRIQGYQIEYRDILNDSSWNAYD 5250
QY 648 -----LNIVINGKSTY----- 662
Db 5251 YIIKDTKYQLYNLNG-SEYEFRIKAKNAAGLSKSPSSLRFLKLGKFTVPSPGAPQVT 5309
QY 663 -----GY 664
Db 5310 RVGKNYVDLKWKEPLRDGSGRITGY 5334

RESULT 12
ABB98574
ID ABB98574 standard; protein; 2835 AA.
XX
AC ABB98574;
XX
DT 14-JAN-2003 (first entry)
DE Dextran saccharase, DSRE.
KW Dextran saccharase; enzyme; cytostatic; dermatological; antiseborrheic;
KW DSR-E; glycosyl transferase; dextran; prebiotic; pharmaceutical;
KW microflora regulation; intestinal transit; mineral assimilation;
KW colon cancer; acne; dandruff; body odour.
XX
OS Leuconostoc mesenteroides NRRL B-1299.
XX
FH Key Location/Qualifiers
FT Peptide 1..40
FT Protein /label= Signal_peptide 41..2835
FT Peptide /label= Mature_protein 423..439
FT /label= SEQ_ID_6
```

FT Peptide /note= "This sequence is specifically claimed in Claim 7"  
FT 478..501  
FT /label= SEQ\_ID 7  
FT /note= "This sequence is specifically claimed in Claim 7"  
FT 519..539  
FT /label= SEQ\_ID 8  
FT /note= "This sequence is specifically claimed in Claim 7"  
FT 560..571  
FT /label= SEQ\_ID 9  
FT /note= "This sequence is specifically claimed in Claim 7"  
FT 631..645  
FT /label= SEQ\_ID 10  
FT /note= "This sequence is specifically claimed in Claim 7"  
FT 1014..1021  
FT /label= SEQ\_ID 11  
FT /note= "This sequence is specifically claimed in Claim 7"  
FT 2120..2138  
FT /label= SEQ\_ID 12  
FT /note= "This sequence is specifically claimed in Claim 7"  
FT 2161..2184  
FT /label= SEQ\_ID 13  
FT /note= "This sequence is specifically claimed in Claim 7"  
FT 2202..2214  
FT /label= SEQ\_ID 14  
FT /note= "This sequence is specifically claimed in Claim 7"  
FT 2243..2250  
FT /label= SEQ\_ID 15  
FT /note= "This sequence is specifically claimed in Claim 7"  
FT 2315..2322  
FT /label= SEQ\_ID 16  
FT /note= "This sequence is specifically claimed in Claim 7"  
FT 2689..2696  
FT /label= SEQ\_ID 17  
FT /note= "This sequence is specifically claimed in Claim 7"  
FT 1981..1142  
FT /note= "Catalytic domain, SEQ ID 1. This sequence is specifically claimed in Claim 4"  
FT Domain  
FT FR2822163-A1.  
XX 20-SEP-2002.  
XX 19-DEC-2001; 2001FR-0016495.  
XX 16-MAR-2001; 2001FR-0003631.  
XX (CNRS ) CNRS CENT NAT RECH SCI.  
XX Bozonnet SAM, Remaud SMMC, Willemot RML, Monsan PEF;  
XX WPI; 2002-715213/78.  
XX N-PSDB; ABQ80961, ABQ80962.  
XX New glycosyl transferase enzymes, containing glucan bonding and  
XX catalytic domains and producing alpha-(1-2) branched dextrans, useful  
XX in probiotic, pharmaceutical or cosmetic compositions  
XX Claim 6; Page 65-74; 82pp; French.  
XX The present sequence is a novel dextran saccharase, DSR-E, from  
XX Leuconostoc mesenteroides NRRL B-1299. The saccharase has glycosyl  
XX transferase activity suitable for producing dextrans having alpha(1-2)  
XX branches from sucrose, D-nitrophenyl-alpha-D-glucopyranoside,  
XX alpha-D-glucogalactose, D-nitrophenyl-alpha-D-glucopyranoside or  
XX alpha-D-galactopyranosyl-sucrose. The dextran saccharase is useful in  
XX probiotic, pharmaceutical or cosmetic compositions. The dextrans and  
XX related compounds having alpha(1-2) bonds, produced using DSR-E, may be  
XX involved in signalling/cellular recognition processes in vivo  
XX (specifically in regulation of microflora in the intestines or on the  
XX skin); and are potentially useful for improving intestinal transit,  
XX increasing assimilation of minerals (e.g. calcium and/or magnesium),  
XX preventing cancer of the colon and combating skin problems such as acne,  
XX dandruff and body odour.

XX SQ Sequence 2835 AA;  
Query Match 26.5%; Score 1165.4; DB 23; Length 2835;  
Best Local Similarity 10.8%; Pred. No. 2.2e-08;  
Matches 285; Conservative 160; Mismatches 202; Indels 1986; Gaps 120;  
QY 2 PVV--LADTSSEDA--LNI-----SPKKEVAENKEGK-----30  
DB 196 PVANKADTSSIHDPPLDTNPTPKSANLUSTQKSTDNQVKSTETSHLOEINKTYFL 255  
QY 31 -----ENHSAMB-----38  
DB 256 DDNGVKKNPFAIIDGKLVFPDKTSGELTANAPQVTGKLVNIDNHNAAHDLTDNFTNV 315  
QY 39 -----TSODF-----KEKTVAV-----50  
DB 316 DGYLTANSWVRPKDILKNGTWTPTTAEDFPRLMSWPPDKNTQVAYLQVYQSGMLPDD 375  
QY 51 -----IKEKVS-----KNPVID-----64  
DB 376 VKVSDNDNMSTLTDAAMTVQKNIESRIGVSGKTMWKODMKNKLIDSOAMNIDSESKND 435  
QY 65 -----NNT-SNEAKIKE-----ENSNK--81  
DB 436 HLOGGALLVNDKTPNANSDYRLNRTPTQTGOITDPSKOGYEMULANDVNSNPV 495  
QY 82 -----SQGYTDSF-----V91  
QY 496 QAEQLNMLHYMNIGTIAQNDPTANFDGYRVADVNDVADLQIAGDYFKAAYGTGKTEA 555  
QY 92 NKN-----94  
DB 556 NANNHISILEDWNNDSAYIRAHGNNQLTMDPRAHLAKTALNPLAAQSGLEPLINTSL 615  
QY 95 -----TEN--PKK-----EDKV-----VYI107  
DB 616 VKRGKDATENEAQNVAFIRAHDESEVQVIAQIHKDKINTKSDGLTTPDEIKQAFITYN 675  
QY 108 A-EFK-DKESGEKAKE--LSSLKQT--KVLV-----TYD-----136  
DB 676 ADELKADKEVTAAYNIPASVAVLLTNKDTVPVYVYDGLPSDDGYMSQKSPYDAITSLK 735  
QY 137 -RI-----FNGSAIETP-----DNL-----151  
DB 736 SRIRYVAGGQSMNTYLHECFDPKAKNETKPGQGLTVSRYGKAMTADBLGNSDTRQOIG 795  
QY 152 -----151  
DB 796 LVINNKPFLLNDDBOIVLWNGAAHKQOAYRPLMLTTKSGIQTIDKXAGAVVYTTNDAG 855  
QY 152 ---DKI---KQIEGI-----SSVERAQKVQ-----170  
DB 856 LIFKSDMVGVSNPQSGYFAAWVPVGASDQDARTOSSQSETKGDVYHSAALDSNVI 915  
QY 171 ---PMN-----174  
DB 916 YEGFSNFQAMPEKNDFTTNVIAQNAKLFKDLGITSELABQVSSSTDSNFSLDVIONGY 975  
QY 175 -----HARKEIGVEAIDYK-----190  
DB 976 AFTDRYDVGVNTPTKYGVQDQLDLSRALHAQ---GIQAINDWPPDQIYNLPGEOIVTAV 1032  
QY 191 -----SINAP-195  
DB 1033 RTNCGKYVDYSVINNTLYDSRTVGGEYQEKFGGLFLDQKKDYPSLFTFKQJSTNQPM 1092  
QY 196 ---F-GKNFQDRCM--VIS-----NI--DNC-----213  
DB 1093 NPDKIKEMSAKYPNGSNIQGRGAWYVLKQMATNQYFNVSNDNPLPKQLIGEXTSGCFI 1152  
QY 214 ---TDYRHRKAMRIDD-----DAKASMFKEKEDLGTGKNTYVLSDKIP---252

Db 1153 TENGKTSFYSTSGYQAKDTFIODGTNNWYFDNAGYMLTGKQNI--HDKNYIF--LPNGV 1207  
 Qy 253 ---HARN---YYN---GGKITVEKYDD---GR---P----- 273  
 Db 1208 ELQDAYLFDGNOEFYFNKAGEQVMQYQDSQONWHYFFENGMAIGLTEVPNADGTHVT 1267  
 Qy 274 -YFDPHGMHIAG-----  
 Db 1268 QYFDANGVOIKGTATKQNNQLRYFDEATGNMNVNSWGQADKSWLYLNAQGVAVTGNQK 1327  
 Qy 290 -DTEQ-----DIK----- 296  
 Db 1328 IDGEYFNADGQKVGNALIDNNGDQRYDGDGKGMVNVNSWGEPLDGSWLYLNDKGIATV 1387  
 Qy 297 ---NFGIDGIAPIAIF-SYKMYSDAGSGFAGDE-----TWP- 330  
 Db 1388 TGRQVINQVNFPGNDG---KQIDAPKLLSD-GSWVYLDKGLITTGAKVINGLNMF 1442  
 Qy 331 ---HAIE---DSI---KHNV---VVSUSSG-----FTG----- 352  
 Db 1443 DKQHOIRKGDASTDANGKRHYDKNDGHLVTNSWGEPLDGSWLYLEEGDAVTGORVIDG 1502  
 Qy 353 ---TGL--VGEK--Y----- 360  
 Db 1503 KTRYFDEDEGKQIKNSLKTLANGDKIYLDGDGVAATGLOHVGDKIMYFDEDEGQVVGKFSV 1562  
 Qy 361 ---W-----QAIRALKAGIPMVVATGNVATSSA----- 386  
 Db 1563 AKDGSWYLYNQDGAAGVSSINGQSLVFDQDGKQVKNVNRNSDGTNNYTGLTGEKLT 1622  
 Qy 387 ---SSW---DL-----VAN-----NHLKM---TDT 403  
 Db 1623 QDFGELPDGSWYLYDAQGHVTVGAQIINGQNLVFKADGQVKGHAYTDQLGHRFYDPDS 1682  
 Qy 404 GN-----VTRTAHE-----DAIVASAKNQ-TVE-----F 428  
 Db 1683 GMLSNRFEQITPGVWAFGADGVAITGQHDINGQKLPFDETGQVQKGSORTIDGTLYSF 1742  
 Qy 429 DK----- 430  
 Db 1743 DSOTGNQKRVQITLLPOAGHYITKNGNDWQYDTNGELAKLRQDSNGKLRYPDLTTGIAQ 1802  
 Qy 431 ---VNIGES---FK-----YR----- 441  
 Db 1803 KQGFVTIGQETVYFSKDHGAQLPMVTEGHYGTITLKQGDQTKTAWVYRDQNNITLKL 1862  
 Qy 442 -NI-G-AFFDKSKITTNE--DGTKAPSKLKFVYI--GKG-----QDQDLIGL 483  
 Db 1863 QNINGTLQFFDP---YTGEQLKGGVAKYDDKLFYFESGKGNLSTVAGDYQDGHVYSQDG 1919  
 Qy 484 R-----GKIAVMD-----RIYTKDLK----- 499  
 Db 1920 QTRYADKQNLVKGLVTNGALQYFDNATGNQIKNQQVIVDGKTYVFDKGNGEVLTNT 1979  
 Qy 500 ---NAP----- 514  
 Db 1980 LDMSTNAFSTKNVAFNHDSSFDHTVDGFLTADTWYRPSILANGTWTWTDSTDKMRPLI 2039  
 Qy 515 VV-----NT-VNYN--RDNWTELPAMGY--BADE-----GTSQVFSIGDD 552  
 Db 2040 TVWPNKVQVNYLNFMKANGLLTTAAQYTLHSDQYDLNQAAQDVQVVAIERIASEHGT 2099  
 Qy 553 G-----VK---LWNM-----INP----- 562  
 Db 2100 WLQKLIFESONNPNPFSVFKQFIWNKDSYHVGGDAMFQGGYLYKGNPLTPTTNSDYQP 2159  
 Qy 563 ---DKTEV-----KENNKEDF--KDK 579  
 Db 2160 GNAFDPLANDVNSNPVQVAENLNLWLYMNFGITAGQDANFDSIRIDAVDFIHTD 2219  
 Qy 580 LEQYV----- 584  
 Db 2220 IQRTYDLRDYQVQOSEAKANQHISLVEAGLDAGTSTIHNDALIESNLREAATLSLTNE 2279

Qy 585 ---P-----ID----- 587  
 Db 2280 PGKNKPLTNMLQDVGGTQLTRDHTQNSTENQATPNYSIIHAHDKQVQKVGAAITDATGA 2339  
 Qy 588 ---MESF-----NS-NKENV-----GDE----- 601  
 Db 2340 DWTNFTDEQLKAGLELFYKQORATNKKNSYNIISIVALMLTNKDTVPRMYYGDMYQDDG 2399  
 Qy 602 ---KBID-----FKF----- 608  
 Db 2400 QYMANKSIYYDALVSLMTARKSYVSGGTMSVDNHGLLKSVRFGKDAMTANDLGTSA 2459  
 Qy 609 ---AP-----DTDK-----ELYKEDI----- 622  
 Db 2460 EGLGVIIIGNDPKQLNDSKVTLDGMAAHKNQYRAVILITTRDGLATFNSDQAPTAMTND 2519  
 Qy 623 ---VPAGST----- 628  
 Db 2520 QGTLTFSNOEINGQDNTQIRGVANPOVSGYLAHVVPVGASDNQDQARTATTENHDGKVL 2579  
 Qy 629 ---SWG-----P----- 632  
 Db 2580 HSNALDSNLIYEGFSNFQPKATTHDELTVVIAKNADVFNWGITSPEMAPQYRSGDH 2639  
 Qy 633 ---RIDLLK-P-----DV-----SAP 645  
 Db 2640 TFLDSTIDNGVAFTRDYDLGFTPTKYTGDLRATTOALHANNQVNVADVVDNOVYNLP 2699  
 Qy 646 GKNIKSTL----- 653  
 Db 2700 QKEVVSATRAGVYGNDDATGFGTQLYVTNSVGGQYQEKYAGQYLEALKAKYPDLFEGKA 2759  
 Qy 654 ---NVIN-GKSTY-----GY 664  
 Db 2760 YDYVYKQVANDGSPNYTSLSHGDRSIPADVAIKQWSAKYMWGNTNVLGNGMGY 2812  
 RESULT 13  
 ABG74786  
 ID ABG74786 standard; Protein; 31267 AA.  
 AC ABG74786;  
 DT 05-JUN-2003 (first entry)  
 DE Human RGS11 protein.  
 KW RGS11; human; screening; cardiant; antianginal; gene therapy;  
 KW heart disorder; cardiac ischaemia; heart failure; angina.  
 OS Homo sapiens.  
 PN WO2002103355-A1.  
 PD 27-DEC-2002.  
 PF 17-JUN-2002; 2002WO-JP06019.  
 PR 18-JUN-2001; 2001JP-0183038.  
 PA (TAKE ) TAKEDA CHEM IND LTD.  
 XX Koyama N, Tanida S, Yamamoto K;  
 XX WPI; 2003-167557/16.  
 XX N-PSDB; ABX13540.  
 PT Screening compounds regulating RGS11 expression and activity for  
 PT prevention and treatment of heart disease -  
 PS Claim 1; Page 59-261; 321pp; Japanese.  
 XX

CC This invention describes a novel method for screening compounds for their  
 CC ability to regulate the activity and expression of human RGS11 and its  
 CC partial peptides and salts, by observing the expression or activity of  
 CC RGS11 in the presence or absence of the test compound. The products of  
 CC the invention have cardiant and antianginal activity and can be used for  
 CC gene therapy. The methods and compositions are useful in the prevention,  
 CC treatment and diagnosis of heart disorders such as cardiac ischaemia,  
 CC heart failure and angina. This sequence represents the human RGS11  
 CC protein described in the disclosure of the invention.

XX Sequence 31267 AA;

Query Match 26.5%; Score 1165.2; DB 24; Length 31267;

Best Local Similarity 8.8%; Pred. No. 0.00025;

Matches 312; Conservative 167; Mismatches 170; Indels 2908; Gaps 139;

```

QY 1 YPVV-----LADT-----SSSED--ALNT--SDKEV----- 23
DB 9448 LPVFDCAISGSEPIVSWYKDKPLKDSPNVQTSFLDNTATLNFKDRSLAGQVSGTA 9507
QY 24 -----AENK-----EKH----- 30
DB 9508 TNPISASSASARLITTEGKNPPFDIRLAPVDAVNGESADEFECHVTGTQPIKVSNAKSR 9567
QY 31 -----EN-TH-----SAMETSOD-----F 43
DB 9568 EIRSGGKQIISYLENSAHLTVLKVKGDSQGYTCVAVNEVGSDCTAQLNIKERLIPPSF 9627
QY 44 KEKTAIVIKKE-----VSKNPVT--DNN-----TSNEAKIK----- 75
DB 9628 TKRLSETEETEGNSFKLEGKRVAGSQPITVAMWKNNIEIQTPSNCEITFKNNTLVLOVRK 9687
QY 76 -----EENSNS-----Q----- 83
DB 9688 AGANDAGLYTCVNSDAGSALCTSSIVIKPKPPVFOHLLPTPTVSGEYVQLSCHVOG 9747
QY 84 -----GDY-----TDS-- 89
DB 9748 SEPRIQWLKAGREIKPSDRCSFSAAGTAVLELDVAKADSGDYVCKASNVAGSDTTKS 9807
QY 90 -----FVN-- 92
DB 9808 KVTIKAEPIQFTKRIQNIIVSEHQATECEVSPDDAIVTWYKGPTELTESQKVFRRDNG 9867
QY 93 ----- 92
DB 9868 RCHWYTHNVTPDDEGVYSVIALRPEGEARSTAEYLTTKEIKLEKRPDIPOSRVP 9927
QY 93 ----- 92
DB 9928 TMPIRAVPEEIPVVAAPRILPLPTPEBKPPPKRIESHESHERKVPKAVPEKKAPPPKV 9987
QY 93 -----KNT-- 95
DB 9988 IKKPVIEKIEKTSRMEEEKVQTVKPEIKPAIPLPAPBPKPBPAGVPKPTPSIEABR 10047
QY 96 -----ENPKEDKVVY-----IAFKD--KES--GEKAIKE--LSLKTAK 130
DB 10048 KRLRPGSGEKPDPDAPFYOLKAVPLKVFVKIKIDIIITSEFVSSAIFECTVSPSTAI 10107
QY 131 VLYTYDRIFNGSAIETP-----DNLDK----- 153
DB 10108 TTMKMD--GSNIESPKGRFIADGDKDKLHIIVQULSDAGEYTCVLRNGKEKTSIAK 10163
QY 154 -----IKQIE-----GT-- 160
DB 10164 LVVEELPVRFVYKTLREEVTVVKGQPLYLSCELNKERDVVWRKDKIIVKEKGRIVPGYIG 10223
QY 161 -----SSVE-----RAQKVQPM----- 172
DB 10224 LMRALTINDADDTAGTYTVTVVENANNLECGSSCVKVEVIRDMLVKPIRDQHVKEKGTAI 10283
QY 173 -----NNH----- 175

```

```

DB 10284 FACIDAKDTPNIKWFPGYDEIPAEBNDKTEILRDGNHLYLKIQAAMPEDIAEVAVEIBGK 10343
QY 176 -----AKK----- 178
DB 10344 RYPAKLTGEEVELELKEPIEDVTIYKESASPDABEISADIPGQWKLKGLLRSPPTCEI 10403
QY 179 -----EIG-----VEE-AIDY--LKSINAP----- 195
DB 10404 KAEQGRFLTHKVKLDAGEVLYOALNATLALTALTYAEIEDLDPVAPLKDVTVPERRAR 10463
QY 196 F-----GKN-----FD-----GRGVNISNIDTGD----- 215
DB 10464 FECVLTREANIVWSKGPDIKSSDKFOIADGKHILVINSQFDDDEGVTAAYEGCKTTS 10523
QY 216 -----YHKAMRT----- 223
DB 10524 ARLPVGTIRLKFMSPLBEDQVKEGETATFVCELSHEKXNVVWPKNDAKLHRTVLISSE 10583
QY 224 -----DD----- 225
DB 10584 GKTHLEKKEVTLDDISOIKAQVKELSTAOQLKVLADPYFTVKLHDKTAVEKDEITLKC 10643
QY 226 -----DAKAMRF-----KEDDLK-----GTDKN-- 244
DB 10644 EVSKDVPV-KWPKDGEIIVSPKYSIKADGRLRIKLIKAKDLKDKGEYVCCCGTDKTKAN 10702
QY 245 -----Y-----W-----LSDKI 251
DB 10703 VTVEARLIKVEKPLVGVFVPGETAHFEILESPDHQWMLKQGPLTASPDCEIIEBGK 10762
QY 252 PHAFNYVN--G-GKTIYE-----KYDD----- 270
DB 10763 KHLILHNCQLGTMGEVSPQAAKSAANLKVKELPLIFITPLSDVKVFEKDEAFCEV 10822
QY 271 -----GRDYF--D--PHGM-----HIAG--I 285
DB 10823 SREPTFRMLKGTQEIITGDDFFELIKDGTGKSMVYKSAAFDEAKYMEAEADKHTSGKI 10882
QY 286 LAG-----ND-----TEOD-- 294
DB 10883 IEGIRLKEPLTKDVTAKESAVFVELSHDNIRVKVKFKDQRLHTRSVSMODEGKTH 10942
QY 295 -----IKANGT----- 301
DB 10943 SITFKDISIDTQIRVEAMGMSSEAKLTVLEGDPYFTGKLQDYTVGEKDEVIILOCEISK 11002
QY 302 -----DG--IAP--NA-----QIPSYKM-----YS-DAGSGFAGDETM-- 329
DB 11003 ADAPVKMFKDQKEIKPKSNANVIAKADGKKRMILKALKKSDIGQYTTCCDGT--DKTSGK 11058
QY 330 -----FPAI-----EDSI----- 337
DB 11059 LDIEDREIKLVPRPHSVVEVMEETARFETIESEDIDHANWKLKGBALLQTPDCEIKEEGK 11118
QY 338 -----KRN-----VDV-----VSVSG-- 349
DB 11119 IHSVLHNCRLDQGTGVDFQANYSANHLKVKRPVIGLARPLKDYTVTJAGETAFFDEL 11178
QY 350 ----- 349
DB 11179 SYEDI PVEMWYKGLKLEPSDKVPRSEKVTTLRLRDVLYLEDAEVOULTAKDPKTHANLF 11238
QY 350 -----FT-----GT-GLVGEKY----- 360
DB 11239 VKBPPEFTKPLEDQTVBEGATAVLECEVSENAKVWFKKGTETLSKKYVEIVADGRVR 11298
QY 361 ----- 360
DB 11299 KLVIHDTCPEDIKTYTGDADKDKTSCNLNVVPRPVEFLRPLJDLQVBEKEWAFECESLR 11358
QY 361 -----W-----QAIRAL-----RKAGIPNV- 375

```

Db 11359 ENAKVFKDGAIEIKKKYDIISKGAVRIILVINKLLDDEAEYSCVETARTSGLTVL 11418  
 Qy 376 ---VATGNYATSASSS---W---D----- 391  
 Db 11419 EEEAVFTKNLANIEVSETDTIKLVCEVSKPGAIEVIWYGDIEIETGRYEILTEGRKRL 11478  
 Qy 392 LVANNHLKMTDGNV---TRT---AAHE----- 413  
 Db 11479 VIQNAHLE--DAGNYNCLPSSRTDGKVKVHELAAAEFISKPNLEILGEKAEFVCSISK 11536  
 Qy 414 ---DAIA-----VASA----- 421  
 Db 11537 ESFPVQWKDDTLESQKYDVIADGKKRVLVVKDATLQDMGTYYVMVGAARAAHLTVI 11596  
 Qy 422 ---KNQTEPD-KVNIAGESFKY-RNIGAPFDKSK----- 451  
 Db 11597 EKLRIWPLKTRVKEQEVVENCEVNTGAKAKWFRNEEAFDSSKYIILQKDLVYTLR 11656  
 Qy 452 ---ITTND----- 457  
 Db 11657 IRDAHLDDQANNVSLTNHRGENVKSAAANLIVEEEDLRIVEPLKDIEIWEKKSVTFWCKV 11716  
 Qy 458 ---GKAPSKLKFEV-----IKGQD----- 475  
 Db 11717 NRLNVLTKWTKNGEVPFNDNRYSYRVKYLKHTIKDCGFPDEGEVIVTAGQDKSVAELL 11776  
 Qy 476 --- 475  
 Db 11777 IIEAPTEFVHELDQTVTFDDAVFSQLSREKANVKYRNREIKEGKKYKFEKDGSIH 11836  
 Qy 476 ---ODLI---GLD----- 482  
 Db 11837 RLIIKDCRLDDECEYACGVEDRKSRARLFEVIEIRPPDIIIEAPGAOVVFLAELNK 11896  
 Qy 483 ---LR-----GKI---AVMD-----RIYTKD----- 497  
 Db 11897 DKVEVQWLENNVVVQGDQKQWSEGIHRLQICDIKPRDQGEYRFAIKDKARAKLELA 11956  
 Qy 498 ---LKNA-----FK----- 503  
 Db 11957 AAPKIKTADODLVVDVGKPLTMVVPYDAYPKAEAEWFKENEPLSTKTIDTTAEQTSFRIL 12016  
 Qy 504 ---KAMDKG----- 509  
 Db 12017 EAKKGDGKRYKIVLQNKHGKASGFNLKVIDVPGVPRNLEVTEFTDGEVSLAWEEPDTG 12076  
 Qy 510 --- 509  
 Db 12077 GSKIIYVVERRDIKRTWVLATDRAESCEFTVTGLQGGVEYLFPRVSARNRVGTGEPVE 12136  
 Qy 510 ---ARA---TMVN-----TVNY----- 521  
 Db 12137 TDNPEARSKYDVPGLPNTITDNRFGVSLTWPEPEYDGGAEITNYVIELDKTSIRW 12196  
 Qy 522 ---Y---NR----- 524  
 Db 12197 DTAMTVRAEDLSATVTDVVEGGEYSFRVRAQNRIGVGKPSAATPFVKVADPIERPSPPVN 12256  
 Qy 525 ---DNW----- 527  
 Db 12257 LTSSDOTQSSVQLKWEPLPKDGGSPILGYIIBERCEBGKNWIRCNMKLVPELTYKVTGLE 12316  
 Qy 528 ---TEL-----P 531  
 Db 12317 KGNKYLRYSAENKAGVSPSEILGPLTADAFVPTMDLSAFKQGLEVIVPNPITILVP 12376  
 Qy 532 ANGYE-----ADEG-----TKSQVFSIS 549  
 Db 12377 STGYPRPTATWCFGDKVLETGDRVKMKTLISAYABLVISPSRSDKGIYTLKLENRVKTLIS 12436  
 Qy 550 G-----DDG-----VKLWNMI 560  
 Db 12437 GEIDNVNVIARPSAPKELKFGDITKDSVHLTWEPDPPDDGSGPLTGYVVEKREVSRTKTVK 12496

Qy 561 ---N-----PDKKTEVK-- 569  
 Db 12497 MDFVTDLFTVVDLVQCKEYLFVKARNKCGPGEPAYVDEPNMSTPATVDPDPENVKWR 12556  
 Qy 570 -RNNKEDF-----RDK-----LEQ--YY 584  
 Db 12557 DRTANSIFLTWDPKPNKDGSRIGYIVVERCPRGSDKWVACGEPVAETKMEVTGLEGKWY 12616  
 Qy 585 -----P-----IDM-----ESP----- 591  
 Db 12617 AYRVKALNRQAGSKPSRPTBIEIOAVDTQEAPEIFLDVKLLAGLVKAGTKIELPATVTGK 12676  
 Qy 592 -----NSNKP----- 596  
 Db 12677 PEPKITWTKADMILKQDKRITTIENVPKSKTVTIIVDSKRSDTGTVIIIEAVNVCGRATAVVE 12736  
 Qy 597 -NVGDE--KEIDFKFA-----P-D-----TDKELY----- 617  
 Db 12737 VNVLDKPGPPAAFDITDVNTESCLLTWNPPRDDGSGKITNVVVERRATDSEVWHKLSSTV 12796  
 Qy 618 -----KEDI-----IV-----PAGSTS----- 629  
 Db 12797 KDTNFKATKLIPNKEYIFRVAENMYGVGEVQASPIAKYQFDPGPPTRLEPSDITKD 12856  
 Qy 630 -----WGPRIDLLKPD----- 641  
 Db 12857 AVTLTWCEPDDGSGSPITGYWVERLD---PDTKWVRCNMKPVKDTTYRVKGLTNKKKY 12912  
 Qy 642 -----VSAPGKNIKT-----LNVIN-----GKS 660  
 Db 12913 RFRVLAENLAGPGKPSKSTELIKPIDPPSPRWLEVINITKNTADLKWTVPKDGGS 12972  
 Qy 661 ---TY-----GY 664  
 Db 12973 PITNYIVEKRDVRRKGW 12989  
 RESULT 14  
 ABB66417  
 ID ABB66417 standard; Protein; 4345 AA.  
 XX  
 AC ABB66417;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 26043.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 WIPI: 2001-656860/75.  
 DR N-PSDB; ABL10520.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX





Db 1940 LWNVTGLGTGVHMQNITDPEYSYHVFTPLAGPENGFIQVLLKKEFVFQHLGKMSNY 1999  
Qy 453 -----TNNEDGT-----KAPSKUK-----FVY----- 469  
Db 2000 KLGKINGEPKSDLVNLQSGNMELEMLYDADFKNLNAETDYKPDADBEYFSYFTNFQVD 2059  
Qy 470 -----TCKGQODLIGLDL-----RGKIAMVDRIYTKDL----- 498  
Db 2060 TLVWPTIVGNVDIIEIDFYLVGHVHVELPQGVKFEKRLHYPDYINVHNLTLVTPPPAVA 2119  
Qy 499 -----KNAF-----KKAMD----- 507  
Db 2120 KNIKSIHVHDLNFAFVERVKFIVNDDKNTQELGFVFNFTALQDNVKKPKAHVDQVTL 2179  
Qy 508 -----KG----- 509  
Db 2180 LTPYEMLHIYVGHIELDDNAYKGNISAVTAHTLSMAASIENEDNFLTSGVGLQETD 2239  
Qy 510 -----ARAI-----MVVN----- 517  
Db 2240 AIPHYGQVYFKKDFSAVDKAIDIRFEVTDNGLNQLHISTDWTDPSPYVNAVGRITKT 2299  
Qy 518 -----TVNYNRDN-----WTBL----- 530  
Db 2300 MLPLQWASTVLVIQONPHLNFDLNLLSQGSTAYGARANKKXDVFNIEWTPMKNFR 2359  
Qy 531 ----- 530  
Db 2360 NISMHGTAIRSPDGRYDVSGFLYRNMATYEVTVGAVRMTNSLPIDVVLVQPKAGGRDG 2419  
Qy 531 -----PAM-----CY-----EAD-----EGTKSQV-----F 546  
Db 2420 VIELNIHEAGPKIRFSAIEDGKMKQMSGYSVSKTNGAMDVSVLVESTEPEIARINF 2479  
Qy 547 -----SISGD-----DGVKL----- 556  
Db 2480 YGNLSPNSEGLVGLDLSTPWLKALGIDTVHLHSDVGLFKGGHIVGEYKIGQYGRGSC 2539  
Qy 557 ----- 559  
Db 2540 LWSWILAEDMQLVLENLERNPAKPRIVHASAKYQNPQGTFTQOAGGRLSVDSKNLWDV 2599  
Qy 560 -----INPDKTE----- 567  
Db 2600 NGSAEYKSVDDPKFRVITALPLPVGDRHOLSAYQGNVISQGFNPNDFVLEASYESFEAQ 2659  
Qy 568 -----VKR-----NN-----KEDFKDKLEQ----- 582  
Db 2660 NKLLSRISYKNATNNLKGILGHVEWGKIQNLVVEGDFELLHKQAGREFSAKIITPKPKN 2719  
Qy 583 -----Y-----YP-----IDME-----SPNSNKP-----NVG 599  
Db 2720 EHTFALTGSYDLEKSGHHNVGSLDPASRRITDLDSVSSLSNMHGIFNSTLPTFLNV- 2778  
Qy 600 DEKEDFKFA-----P-----P-----DTD----- 613  
Db 2779 SWLKTDFTNTNGKSRYRCRCFWPDQTYAFKLNSNYDSSNFHNLNGVIEIPLAT 2838  
Qy 614 -----KELYKE-----DIIV----- 623  
Db 2839 RHRADIVYQKRRNQDAGNVVYNEKQVLDGKYKRLQKAKAPIYKETTDISLENEVKP 2898  
Qy 624 -----PAGS----- 627  
Db 2899 LGHFVSTRDASDPAGSQDKHIEIYELNNTQFNLTGELHSRATLKAQDFKVAIHPNR 2958  
Qy 628 -----TSW----- 630  
Db 2959 AVLSTKYEDVSEFVVRHHSKLESETAWIGYNLELGNFSKVGNESQSFALEIFPKNML 3018  
Qy 631 -----GPR----- 633  
Db 3019 SSSGQYYMTDTNFNSDLSFQWLGNGYDQPKIIHNSNLQWKAEPHLRGDREHRTIALTVAH 3078

Qy 634 -----IDLLKPP-----VSAP 645  
Db 3079 PLLSKDINCKATYYRGLRDLRLTHLTIDYSEYDQOLIEGLQALTDYRSELGHTNVTFTHY 3138  
Qy 646 GKNIKSTLNV-ING-----KST-----YG 663  
Db 3139 GKHTASELDVOLNGTLAAMNSYYKTESTAHYKRDIFPARYG 3179  
RESULT 15  
AAU05396  
ID AAU05396 standard; Protein; 26926 AA.  
XX AC AAU05396;  
XX DT 24-OCT-2001 (first entry)  
XX DE Human titin (connectin) protein sequence.  
XX KW Human; titin; connectin; pickwick mutation; cardiac specific exon; N2B;  
KW titin-related disease; zebrafish; heart failure; heart disease.  
XX OS Homo sapiens.  
XX PN WO200151666-A1.  
XX PD 19-JUL-2001.  
XX PF 12-JAN-2001; 2001WO-US01212.  
XX PR 12-JAN-2000; 2000US-0175787.  
XX PA (GEHO ) GEN HOSPITAL CORP.  
XX PI Fishman MC;  
XX DR WPI; 2001-451869/48.  
XX DR N-PSDB; AAS05390.  
PT Determining if a subject has or is at risk of developing a  
PT titin-related disease or condition, particularly heart failures,  
PT comprises detecting the presence of a mutation in the titin gene -  
XX Disclosure; Page 57-111; 114pp; English.  
XX The present sequence representing human titin (also known as connectin)  
is described in an invention relating to a novel method for determining  
CC whether a subject has or is at risk of developing a titin-related  
CC disease or condition. The method comprises analysing a nucleic acid of  
CC a sample from the subject and detecting the presence of a mutation  
CC (e.g. the pickwick mutation in the cardiac specific exon N2B) in the  
CC titin gene, which indicates that the subject has or is at risk of  
CC developing a titin-related disease. The zebrafish which has a phenotype  
CC similar to mammalian heart failure is used as a model. The method is  
CC useful for detecting an increased likelihood of heart disease, such as  
CC heart failure, in a patient, so that appropriate intervention can be  
CC instituted before any symptoms occur. The method may also be used to  
CC facilitate determination of etiology of an existing heart condition,  
CC such as heart failure, to identify compounds that can be used to treat  
CC or prevent heart conditions, in prenatal genetic screening, e.g. to  
CC identify parents who may be carriers of a recessive titin mutation.  
CC Compounds identified using the methods may be used to treat patients  
CC that have or are at risk of developing heart disease, e.g. heart  
CC failure.  
XX Sequence 26926 AA;  
SQ

Query Match 26.4%; Score 1161.5; DB 22; Length 26926;  
Best Local Similarity 8.0%; Pred. No. 0.00016;  
Matches 323; Conservative 169; Mismatches 164; Indels 3365; Gaps 140;  
Qy 2 P---VVLAD-----TSSSE 12

```

Db 3809 PETQAVLSDTEKIFPSAMSIEQINSLVEPLKTLLEBEGNVPOSSIEPMMHSYLLSYAE 3868
Qy 13 DAL-----NISD-----KEKVAENK-----EKH-ENI----- 33
Db 3869 EVLSKTEKTVSDTNRQVTLQKQEOAGSALLISQSLAGHVESLOSPPVMIISQVNEPLV 3928
Qy 34 ---HSAME-----TSQD-----FK---EKTAVIKER----- 54
Db 3929 PSEHSCTEGKILIESANPLENAGODSAVRIEKGSLAFPLALEKQVLLKEHSDNVVM 3988
Qy 55 ----- 60
Db 3989 PPDOIIESKREPVAIKVQEOVRDLSESLSGIPEEQRNLKIQICRALQAQVASEQ 4048
Qy 61 P-----VIDNNTSN----- 69
Db 4049 PGLPSEWLRNIEKVEVEAVNITQEPRIIMCVLTSAKSVTEVTIITIEDVPOMANLKM 4108
Qy 70 -----BEAKI----- 74
Db 4109 ELRDALCAIIEEIDILTAEGPRIQOGAKTSLQEMDSFGSQKVEPITTEBEVESKYLIS 4168
Qy 75 -----KEENSNS----- 82
Db 4169 TEEVYFNVQSRVKYLDATPVTKGVA SAVSDEKODESLKPESEKESSSSGTEEVATV 4228
Qy 83 -----OGD---YTDSFVN-----KNTENPKKE--- 101
Db 4229 KIOAEGGLIKEDGEMHTPLVDIVSEGDIVHLTTSITNAKEVNMVFNKLVPSDEKFK 4288
Qy 102 -----DKVV-----YIAE----- 109
Db 4289 CLQDONTTVLVIDKNTEDHOGCEYCEALNDSGKATSAKLTVGRAPVIRKRIEPLV 4348
Qy 110 -----FK-----DKES-----GE- 117
Db 4349 ALGHAKTECEIQSAPNVRFQWFKAGREIYESDKSIRSKYISSLEILRTQVNDGCGYT 4408
Qy 118 -KAIKELSSLKTKVL----- 132
Db 4409 CKASNEGVSCTALTLTVPOGEEKVRKLLPERKPEKEEVLKSLVRKPEEPEKVE 4468
Qy 133 ----- 132
Db 4469 PKULEKVKKPAVPEPPPKVEVEVPTVTKRERKIPEPTKVPEIKPAIPLPAPPEPKP 4528
Qy 133 ----- 132
Db 4529 EAEVKTIKPPVEPEPTPIAPVTVPVVGKKAEMAKAPKEBAKPKGPIKVPKTPPIE 4588
Qy 133 -----YTY-----D-----RIF----- 139
Db 4589 AERRKLREGSGEKEPDEAPFTYQLKAVPLKFVEIKDIITLSESEFVSSAIFECLEVPS 4648
Qy 140 -----NSALETTP-----DMLDK----- 153
Db 4649 TAITTWMKOGSVIRSESPKGRFIADGKRKLLIIVOLSDAGEYTCVLRGNKEKTSTAKL 4708
Qy 154 -----IKOIE-----GI--- 160
Db 4709 VVEELPVAFVKTLEBEVTVVKGQPLYLSCELNKERDVMVRKDGKIVKEPGRIVGCVGL 4768
Qy 161 -----SSVE-----BAQKVQPM--- 172
Db 4769 MRALTINDADDDTAGTYTATVENANNLECSQVKEVIRDLWLKPIRDQHYKPKGTAF 4828
Qy 173 -----MNH----- 175
Db 4829 ACDIAKDPNITKWFQGYDEIPAEPVDKTEILRDGNHLTKIKNAMPEIDIAEVAVEBGR 4888
Qy 176 -----ARK----- 178

```

```

Db 4889 YPAKTLTGEREVELLKPIEDVTIYEKESASFDAISEADIPGQWKLKCELLRSPTEIK 4948
Qy 179 -EIG-----VEE-AIDY---LKSINAP-----F 196
Db 4949 AEGGRFLTLHKVLLDQAGEVLYQALNAITTAITLTVAEIEIDFVPLKDVTPERRQARF 5008
Qy 197 -----GKN-----F---DGRGVNISNIDTGD- 215
Db 5009 ECVLTREANVIWMSGPPIIKSSDKFDIIADQCKHILVINSQFDEGVYTAVEGKKTSA 5068
Qy 216 -----YRHKAMRI----- 223
Db 5069 RLFTVGIKFKMSPLEDOQVAGEATFVCELSHEKKNVWPKNDAKLHSTRVLISSEG 5128
Qy 224 -----DD----- 225
Db 5129 KTHKLEKEVTLDDISQIKAOVKELSTAOULKLEADPYTVKLHDKTAVKEITLKE 5188
Qy 226 -----DAKASMRP-----KKEDLK-----GTDKN--- 244
Db 5189 VSKDVPV-KWFKQDEEIVSPKYSIKADGLRIKIKADLKDQGEVYCCGCTDKTANV 5247
Qy 245 -----Y-----W-----LSDKIP 252
Db 5248 TVEARLIEVEKPLYGVEFVGETAFHEIELSEPPVHGQWKLKQPLTASPCEIIEBGK 5307
Qy 253 HAFNYVN---G---KTIYE-----KYD- 270
Db 5308 HILILHQCQGMTEVSEFQOANAKSAANLKVKELPLIFITPLSVKVPKEDKAEFEVVS 5367
Qy 271 -----GRDYF---D---PHQM-----HIAG--IL 286
Db 5368 REPRTFWMKCTOETIDDRFELIKDGTKSMVKSAAFEDEAKYMEAEBKHSKULI 5427
Qy 287 AG-----ND-----TEOD----- 294
Db 5428 EGIRLKEFLPLKDVTAKEGSAVFTVELSHDNIRVKWFKNDQRLHTRSVMQDEGKTHS 5487
Qy 295 -----IKNPGCI----- 301
Db 5488 ITFKDLISDTSQIRVAMGNSSEAKLTVEGDYFTGKLQDLYGVEDEVILQCEISKA 5547
Qy 302 -----DG--IAP--NA-----QIFSKM-----YS-DAGSGFAGDETM--- 329
Db 5548 DAPVKRDKGEIETPSKNAVKTGDKKMLIKALSDIGQYTCDCGT---DKTSGKL 5603
Qy 330 -----FPAI-----EDSI----- 337
Db 5604 DIEDREIKLVRLHSHVEVMEETETARFETEISEDDIHANMWKLGEALLQTPDCEIKESGI 5663
Qy 338 -----KKN-----VDV-----VSYSG----- 349
Db 5664 HSLVLHNCRLDQGTGVFOANVKSAAHLRVKPRVIGLRLKDVLTAGETATFDCELS 5723
Qy 350 ----- 349
Db 5724 YEDIPVEWYLLKGLLEPSDKVPRSEGKATTLRLDKVLEBAGEVOLTAKDFKTHANLFV 5783
Qy 350 -----FT-----GT-GLVGEKY----- 360
Db 5784 KEPVEFTKPLEDOQVGEATVAVECEVSRENAKVWFKNGKTEILSKSKEYIVADGRVRK 5843
Qy 361 ----- 360
Db 5844 LVHIDCTPEDIKTYTCDAKQFKTSCNLNVPPHVEFLRPLTDLQVREKEMARFECELSRE 5903
Qy 361 -----W-----QAIRALKAGIPW-- 375
Db 5904 NAKVWFKDGAIEIKKGKYYDIISKGAVRILVINKCLLDDENAEVSCVRIARTASGMLTVLE 5963
Qy 376 ---VATGNVATSASSS-----W--D-----L 392
Db 5964 BEAVFTKMLANIEVSEFTDIKLVCEVSGPAEVIWYKGBEIIETGRVEILTEGRKRILV 6023

```

```

QY 393 VANNHLMKTDTCNV-----TRT-----AAHE----- 413
Db 6024 IONAHLE--DAGNYNCRPLSSRTDGVKVKVHELAABEFISKPONLEILEGEKAFFVCSISKE 6081
QY 414 -----DATA-----VASA----- 421
Db 6082 SFPQWKRDDKLTESGDKVDVTDGKKRVLVVKDATLQDMGTYVVMVGAARAAHLTVIE 6141
QY 422 -----KNQTVFED-KYNIGESPXY-RNIGAPFDKSK----- 451
Db 6142 KLRIVVPLKDRVKEQEVVFNCEVNTGAKAKWFRNEBEAIPDSSKYIILQKDLVYTLRI 6201
QY 452 -----ITTNE----- 457
Db 6202 RDAHLDQANYVSLNTHRGENVKSAANLIVEEDLRIVEPLKDIETMEKKSVTFWCKVN 6261
QY 458 -----GTAKPSKLFVY-----IGKQD----- 475
Db 6262 RLNVTLKWTNGEEVFPDNRVSRYVDKYKMLTIKDCGPPDEGEIYTAGQDKSVABELLI 6321
QY 476 ----- 475
Db 6322 IEAPTEFVEHLEDQVTBFDVAVFSCLSREKANVKYRNGREIKEGKYPKEDGSIHR 6381
QY 476 -----ODLI-----GLD----- 482
Db 6382 LIIKCRLLDCEYACGVEDRKSARLFVEEIPVEIIRPPQDILEAPGADVVFVLAELNKO 6441
QY 483 -----LR-----GKI-----AVMD-----RYTKD----- 497
Db 6442 KVEVOWLRNMVVOGDKHOMSEBGIHLRLOICDIKPRDQGEYRFIAKDKARAKLELAA 6501
QY 498 -----LKNA-----FK----- 503
Db 6502 APKIKTADQDLVDVGKPLTMVVPYDAYPKAEAEWFKENEPLSTKTIDTTABQTSFRILE 6561
QY 504 -KAMDKG----- 509
Db 6562 ARKGDGGRYKIVLQNKHGKAEFINLKVIDVPGPVNLEVTFDFGEVSLAWEEPLTDGG 6621
QY 510 ----- 509
Db 6622 SKIIGVYVVERDIKRTWVLATDRAESCEFTVTGLQKGVEYLFVRSARNRVGTGEVET 6681
QY 510 -----AR-----AIMVN-----TVNY----- 521
Db 6682 DNPVEARSKYDVGPPPLNVTITDNRFRGVSILTWEPPDYDGAETINYVIELRDKTSIRWD 6741
QY 522 -----Y-----NR----- 524
Db 6742 TMTVRAEDLSATVTDVVBEGQESYFRVRAQNRIGVGKPSAATPFVKVADPIERPSPVNL 6801
QY 525 -----DNW-----TEL----- 530
Db 6802 TSSDOTQSSVQLKWBEPPLKDGSPILGYIIERCEBGNWIRCNWKLVPETYKYVTGLEK 6861
QY 531 ----- 532
Db 6862 GNKYLVRSAENKAGVSDPSEILGPLTADDAFVEPTMDLSAFKDGLEIVVNPNTILVPS 6921
QY 533 MGYE-----ADEG-----TKSQVPSISG 550
Db 6922 TGYPRPTATWCFQDKVLETGDRVNMKTL SAYAELVISPSRSDKGIYTLKLENRVKTIISG 6981
QY 551 -----DDG----- 559
Db 6982 EIDVNVIAEPSAPKELKFGDITKDSVHLTWEPDDDDGSPLTGCVVVEKREVSARKTWM 7041
QY 560 -----IN-----PDKKTEVK----- 569
Db 7042 DFVTDLEFTVPLVQGEYLFKVCARNKCGPGEPAVVDPEVNMSTPATVDPDENVKWRD 7101

```

```

QY 570 RNNKEDF-----KDK-----LEO--YY- 584
Db 7102 RTANSIFLTWDPKNDGSGRIKGYIVERCPGSDKWACGEPVAETKMEVTLGEEGKYA 7161
QY 585 -----P-----IDM-----BSF----- 591
Db 7162 YRVKTLNRQAGSKPSRPTETIQAVDTOBAPEIFLDVKLLAGLTVKAGTKIELPATVTKP 7221
QY 592 -----NSNK----- 596
Db 7222 EPKITWTKADMILKODKRITITENVPKSTVTIVDSKRSDDTYIIEAVNVCGRATAVVEV 7281
QY 597 NVGDB-----K 602
Db 7282 NVLDKPGPPAAPADITDVTNESCLLTWNPPRDDGGSKITVYVVERRATDSEVWHKLSSTVK 7341
QY 603 BDPK----- 607
Db 7342 DTNFKATKLI PNKEYIFRVAENMYGAGEPVQASPIAKYQFDPGPPPTLRLEPSDITKDA 7401
QY 608 -----FAPDTPK----- 614
Db 7402 VLTWCEPDDDDGSGPITGYWVERLDPDTKWVRCKNMPVKDITTVRVKGLTNKKKYRFRVL 7461
QY 615 -----ELYKEDIIVP----- 624
Db 7462 AENLAGCPKPSKSTEPILIKDIPDPWPDPKPTVKDVGKTSVRLNWTYPBHDGGAKIESY 7521
QY 625 -----AGTSW-----G----- 631
Db 7522 VIEMLTGTDDEWVRVAEGVPTTQHLLPGLMEQEYSFRVRAVNKAGESESPSPVLCR 7581
QY 632 -----PR-----IDLLK 639
Db 7582 EKLPPSPPRWLEVINIITKTADLKWTVPEKDGSGSPITNYIVEKRDVRRKGMQTVDTTVK 7641
QY 640 -----P-----DVS-- 643
Db 7642 DTKCTVTPLTBEGSLYVFRVAENAIGQSDYTEIEDSVLAKDTFTTGPFPYALAVVDVTKR 7701
QY 644 -----APGKN-----IKS-----TLNVI----- 656
Db 7702 HVDLKWEPKNDGGRPIQRYVIEKRLGTRWVKAGTAGPCDNFRVTDVIEGTEVQFQV 7761
QY 657 -----NGKS--TVG 663
Db 7762 RAENAGVGHPSPEPTIELSIEDPTSPSPPLDLHLVTDAGRKHIAIAWKPPKNGGSPIG 7821
QY 664 Y 664
Db 7822 Y 7822

```

Search completed: October 9, 2003, 15:50:54  
Job time : 127 secs

RESULT 1	
AAB48342	
ID	AAB48342 standard; Protein; 664 AA.
XX	
XX	
AC	AAB48342;
XX	
DT	20-APR-2001 (first entry)
XX	
DE	S. pneumoniae Spi28 polypeptide.
XX	
KW	Immunogenic; Spi28; Spi30; pneumococcal; otitis media; nasopharyngeal
KW	bronchial; lung; blood; infection; immune response; immunotherapy;
KW	antibacterial; auditory; vaccine.
XX	
OS	Streptococcus pneumoniae.
XX	
PN	MO200076540-A2.
XX	
PD	21-DEC-2000.
XX	
PF	09-JUN-2000; 2000MO-US15925.
XX	
PR	10-JUN-1999; 99US-0138453.
XX	
PA	(MED1-) MED IMMUNE INC.
XX	
PI	Adamou JE, Choi GH;
XX	
DR	WPI; 2001-112197/12.
XX	
DR	N-PSDB; AAC84741.
XX	
FT	New vaccines comprising Spi28 or Spi30 polypeptides, for treating and



QY 181 GVEBAIDYLSKINAPFGNPDGRGVISNIDTGTDRHKAMRIDDAKASNRFKKEDLKG 240  
DB 181 GVEBAIDYLSKINAPFGNPDGRGVISNIDTGTDRHKAMRIDDAKASNRFKKEDLKG 240  
QY 241 TDKNYWSLSDKIPHAFNYYNGGKITVEKDDGRDYPDPHGMHIIAGILAGNDTEODIKPNFG 300  
DB 241 TDKNYWSLSDKIPHAFNYYNGGKITVEKDDGRDYPDPHGMHIIAGILAGNDTEODIKPNFG 300  
QY 301 IDGIAIPNAQIFSYKMYSDAGSGFAGDETFMFAIEDSIKHNVDVVSVSSGFTGTGLVGEKY 360  
DB 301 IDGIAIPNAQIFSYKMYSDAGSGFAGDETFMFAIEDSIKHNVDVVSVSSGFTGTGLVGEKY 360  
QY 361 WQAIRALRKAGIPMVVATGNVATASASSSSWDLVANNHLMKMTDGTGNVTRTAHEDAIAS 420  
DB 361 WQAIRALRKAGIPMVVATGNVATASASSSSWDLVANNHLMKMTDGTGNVTRTAHEDAIAS 420  
QY 421 AKNOTVEFDKVNIGESFKYRNIGAFDPKSKITTNEDGTAKPSLKLFYIKGODQDOLIG 480  
DB 421 AKNOTVEFDKVNIGESFKYRNIGAFDPKSKITTNEDGTAKPSLKLFYIKGODQDOLIG 480  
QY 481 LDLRGKIAVMDBRITTKDKNAFKKAMDGARAIVVNVVYVNDNMTLPMAGYEADG 540  
DB 481 LDLRGKIAVMDBRITTKDKNAFKKAMDGARAIVVNVVYVNDNMTLPMAGYEADG 540  
QY 541 TKSQVFSISGDDGVKLMNINPDKTEVKRNKKEDEPKDLEQYYPIDMESFNSKNPNVD 600  
DB 541 TKSQVFSISGDDGVKLMNINPDKTEVKRNKKEDEPKDLEQYYPIDMESFNSKNPNVD 600  
QY 601 EKEIDFKFAPDTDKELYKEDIIVPAGSTSWGPRIDLLKPDVSAFGKNIKSTLVNINGS 660  
DB 601 EKEIDFKFAPDTDKELYKEDIIVPAGSTSWGPRIDLLKPDVSAFGKNIKSTLVNINGS 660  
QY 661 TYGY 664  
DB 661 TYGY 664

RESULT 3  
ABU01020  
ID ABU01020 standard; Proteoin; 2140 AA.

XX AC ABU01020;  
XX DT 11-FEB-2003 (first entry)  
XX DE S. pneumoniae type 4 strain protein from coding region #590.  
XX KW Bacterial meningitis; pneumonia; sepsis; otitis media;  
XX KM ear infection; antiinflammatory; antibacterial; immunostimulant;  
XX OS audiotory; respiratory; gene therapy; vaccine.  
XX OS Streptococcus pneumoniae type 4 strain.  
XX PN WQ200277021-A2.  
XX PD 03-OCT-2002.  
XX PF 27-MAR-2002; 2002WO-IB02163.  
XX PR 27-MAR-2001; 2001GB-0007658.  
XX PA (CHIR-) CHIRON SPA.  
XX PA (GENO-) INST GENOMIC RES.  
XX PI Maelignant V, Tettelein H, Fraser C;  
XX DR WPI: 2003-040579/03.  
XX DR N-PSDB; ABX06302.  
XX PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
PT useful as medicaments for treating or preventing a disease or infection  
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media  
PT or ear infection

XX XX  
XX Claim 1; SEQ ID No 1180; 56pp; English.  
PS The invention relates to a protein comprising or having at least 50%  
XX identity to any of the 2469 amino acid sequences, identified in the  
CC specification (available on a computer readable format), or its fragment,  
CC expressed from 2469 of 2489 identified DNA coding regions from the  
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as  
CC AB556454. Also included are an antibody which binds one of the  
CC proteins, treating a patient by administering the protein, DNA or  
CC antibody (in a composition), a kit comprising first and second primers,  
CC which are the nucleic acid cited above or fragments between nucleotides  
CC 8-100 of a sequence not defined in the specification, for amplifying a  
CC target sequence contained within a Streptococcus nucleic acid sequence,  
CC where the first primer is substantially complementary to the target  
CC sequence and the second primer is substantially complementary to the  
CC complement of the target sequence, and where the parts of the primers  
CC having substantial complementarity define the termini of the target  
CC sequence to be amplified, assay comprising contacting a test compound  
CC with the protein, and determining whether the test compound binds to the  
CC protein and a Streptococcus pneumoniae bacterium, where one or more  
CC genes encoding the proteins has been rendered inactive. The proteins,  
CC nucleic acid molecules, antibody and compositions are useful as  
CC medicaments for treating or preventing a disease or infection due to  
CC streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,  
CC sepsis, otitis media or ear infection. They are also useful in developing  
CC vaccines, diagnostics and antibiotics. The methods are useful for  
CC identifying immunodominant proteins. The present sequence is one of  
CC the 2469 proteins expressed by the identified coding regions from the  
CC genomic sequence.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 2140 AA;

Query Match 99.9%; Score 4395.7; DB 24; Length 2140;  
Best Local Similarity 99.6%; Pred. No. 5.6e-74;  
Matches 664; Conservative 0; Mismatches 0; Indels 3; Gaps 2;

QY 1 Y-P--VVLADTSSSEDAINISDEKEVAENKEKHENISAMETSODFEKKTAVIKERBV 57  
DB 19 YAPNEVVLADTSSSEDAINISDEKEVAENKEKHENISAMETSODFEKKTAVIKERBV 78  
QY 58 SKNPVINDNTSNEBAKIKEENSKNSQGDYDTSFVNKKTENPKKEDKRVYIAEFKDSGE 117  
DB 79 SKNPVINDNTSNEBAKIKEENSKNSQGDYDTSFVNKKTENPKKEDKRVYIAEFKDSGE 138  
QY 118 KAIEKLSLKNTKVLVYDRFNGSALETTPDNLDKIKOIEGSSVERAQVQPMNHR 177  
DB 139 KAIEKLSLKNTKVLVYDRFNGSALETTPDNLDKIKOIEGSSVERAQVQPMNHR 198  
QY 178 KEIGVEBAIDYLSKINAPFGNPDGRGVISNIDTGTDRHKAMRIDDAKASNRFKKED 237  
DB 199 KEIGVEBAIDYLSKINAPFGNPDGRGVISNIDTGTDRHKAMRIDDAKASNRFKKED 258  
QY 238 LKGTDKNYWSLSDKIPHAFNYYNGGKITVEKDDGRDYPDPHGMHIIAGILAGNDTEODIKN 297  
DB 259 LKGTDKNYWSLSDKIPHAFNYYNGGKITVEKDDGRDYPDPHGMHIIAGILAGNDTEODIKN 318  
QY 298 FNGIDGIAIPNAQIFSYKMYSDAGSGFAGDETFMFAIEDSIKHNVDVVSVSSGFTGTGLV 357  
DB 319 FNGIDGIAIPNAQIFSYKMYSDAGSGFAGDETFMFAIEDSIKHNVDVVSVSSGFTGTGLV 378  
QY 358 EKYWQAIRALRKAGIPMVVATGNVATASASSSSWDLVANNHLMKMTDGTGNVTRTAHEDAI 417  
DB 379 EKYWQAIRALRKAGIPMVVATGNVATASASSSSWDLVANNHLMKMTDGTGNVTRTAHEDAI 438  
QY 418 VASAKNOTVEFDKVNIGESFKYRNIGAFDPKSKITTNEDGTAKPSLKLFYIKGODQD 477  
DB 439 VASAKNOTVEFDKVNIGESFKYRNIGAFDPKSKITTNEDGTAKPSLKLFYIKGODQD 498

478 LIGDLRGKIAVMDRIYTKDLKNAPFKKAMDKGARAIMVNTVYNNRDNTWTELPAMGYEA 537  
 499 LIGDLRGKIAVMDRIYTKDLKNAPFKKAMDKGARAIMVNTVYNNRDNTWTELPAMGYEA 558  
 538 DEGTQSVFISGGDGVKLMNNINPKKTEVKRNKEDFKDLQYYPIDMESFNSKNPN 597  
 559 DEGTQSVFISGGDGVKLMNNINPKKTEVKRNKEDFKDLQYYPIDMESFNSKNPN 618  
 598 VGDEKEIDFKPAPDTEKELYKEDIIVPAGTSWGPRIIDLLKPDVSAFGKNIKSTLNVIN 657  
 619 VGDEKEIDFKPAPDTEKELYKEDIIVPAGTSWGPRIIDLLKPDVSAFGKNIKSTLNVIN 678  
 658 GKSTYGY 664  
 679 GKSTYGY 685  
 RESULT 4  
 AAW61246  
 ID AAW61246 standard; Protein; 1007 AA.  
 XX  
 AC AAW61246;  
 DT 02-OCT-1998 (first entry)  
 XX  
 DE Streptococcus pneumoniae SP122 protein.  
 XX  
 KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;  
 KW detection; pneumonia; otitis media; meningitis.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 PN WO9818930-A2.  
 XX  
 PD 07-MAY-1998.  
 XX  
 PF 30-OCT-1997; 97WO-US19422.  
 XX  
 PR 31-OCT-1996; 96US-0029960.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Choi GH, Hromockyj A, Johnson LS, Kunsch CA;  
 XX  
 DR WPI; 1998-272224/24.  
 DR N-PSDB; AAV27431.  
 XX  
 PT Nucleic acid encoding antigenic peptide(s) from Streptococcus  
 PT pneumoniae - or their epitope-containing fragments, useful in  
 PT protective or therapeutic vaccines, and for diagnosis  
 XX  
 PS Claim 11; Page 92-93; 118pp; English.  
 XX  
 CC The present sequence represents a protein from Streptococcus pneumoniae.  
 CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein  
 CC can be useful in vaccines for inducing protective antibodies against  
 CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.  
 CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid  
 CC are used to detect Streptococcus infection (by usual hybridisation or  
 CC amplification methods), also for isolating Streptococcus genes or their  
 CC allelic variants. The protein can be used similarly to detect specific  
 CC antibodies in standard immunoassays, especially for diagnosing or  
 CC monitoring infections. Antibodies which bind the protein are used to  
 CC detect corresponding antigens, to purify the protein and for passive  
 CC immunisation (optionally coupled to a toxin). Vaccines are administered,  
 CC e.g. by injection, orally or through the skin, typically at 0.01-1000  
 CC (especially 10-300) mu g/ml per dose.  
 XX  
 SQ Sequence 1007 AA;  
 Query Match 94.8%; Score 4171; DB 19; Length 1007;  
 Best Local Similarity 100.0%; Pred. No. 4.4e-71;  
 Matches 627; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 ETSQDFKKEKTAIVIKEKEVSVSKNPVIDNNTSNEBAKIKEENSKSQGDYTDSEVFNKNTEN 97  
 DB 1 ETSQDFKKEKTAIVIKEKEVSVSKNPVIDNNTSNEBAKIKEENSKSQGDYTDSEVFNKNTEN 60  
 QY 98 PKKEDKVYIAEAFKDKESGEKAIKELSSLKNTKVLYTYDRIENGSAIETTPNLDKIKOI 157  
 DB 61 PKKEDKVYIAEAFKDKESGEKAIKELSSLKNTKVLYTYDRIENGSAIETTPNLDKIKOI 120  
 QY 158 EGISVSERAQVQPMNMNHARKEIGVEEAIDYLSINAPFGKNGFGRGMVINSIDTGTDIR 217  
 DB 121 EGISVSERAQVQPMNMNHARKEIGVEEAIDYLSINAPFGKNGFGRGMVINSIDTGTDIR 180  
 QY 218 HKAMRIDDDAKASMRFKKEDLKGTDKNYLSDKI PHAFNYNGGKITVVEKYDDGRDYFDP 277  
 DB 181 HKAMRIDDDAKASMRFKKEDLKGTDKNYLSDKI PHAFNYNGGKITVVEKYDDGRDYFDP 240  
 QY 278 HGMHIAGILAGNDTEQDIKNFNIGIDGIAPNAQIFSYKMYSDAGSGFAGDETFPHATEDSI 337  
 DB 241 HGMHIAGILAGNDTEQDIKNFNIGIDGIAPNAQIFSYKMYSDAGSGFAGDETFPHATEDSI 300  
 QY 338 KHNVDVSVSSGFTGTGLVGEKYWQAIIRALRKAGIPMVVATGNYATSASSSSWDLVANNH 397  
 DB 301 KHNVDVSVSSGFTGTGLVGEKYWQAIIRALRKAGIPMVVATGNYATSASSSSWDLVANNH 360  
 QY 398 LKMTDTGNTVTRTAHEDAIASAKNQTVFDPKVNIGGESPKYRNIGAFDFKSKITTND 457  
 DB 361 LKMTDTGNTVTRTAHEDAIASAKNQTVFDPKVNIGGESPKYRNIGAFDFKSKITTND 420  
 QY 458 GTKAPSKLKFVYIGKGQDGLIGDLRGKIAVMDRIYTKDLKNAPFKKAMDKGARAIMVYN 517  
 DB 421 GTKAPSKLKFVYIGKGQDGLIGDLRGKIAVMDRIYTKDLKNAPFKKAMDKGARAIMVYN 480  
 QY 518 TVNYYNRDNWTELPAMGYEADGTSQVFSISGDDGVKLMNNINPDKTEVKRNKEDFK 577  
 DB 481 TVNYYNRDNWTELPAMGYEADGTSQVFSISGDDGVKLMNNINPDKTEVKRNKEDFK 540  
 QY 578 DKLEQYYPIDMESFNSKNPNVGDEKEIDFKFAPDTEKELYKEDIIVPAGTSWGPRIIDLL 637  
 DB 541 DKLEQYYPIDMESFNSKNPNVGDEKEIDFKFAPDTEKELYKEDIIVPAGTSWGPRIIDLL 600  
 QY 638 LKPDVSAFGKNIKSTLNINVGKSTYGY 664  
 DB 601 LKPDVSAFGKNIKSTLNINVGKSTYGY 627  
 RESULT 5  
 ABP54664  
 ID ABP54664 standard; Protein; 1007 AA.  
 XX  
 AC ABP54664;  
 DT 04-SEP-2002 (first entry)  
 XX  
 DE S. pneumoniae SP122 protein sequence SEQ ID NO:216.  
 KW Streptococcus pneumoniae; epitope; vaccine; antigenic protein;  
 KW antibacterial; Streptococcal infection; detection.  
 OS Streptococcus pneumoniae.  
 PN US2002061545-A1.  
 PD 23-MAY-2002.  
 PF 22-JAN-2001; 2001US-0765272.  
 XX  
 PR 30-OCT-1997; 97US-0961083.  
 PA (CHOI/) CHOI G H.  
 PA (KUNS/) KUNSCH C A.  
 PA (BARA/) BARASH S C.  
 PA (DILL/) DILLON P J.



PA (DOUG/) DOUGHERTY B.  
 PA (FANN/) FANNON M R.  
 PA (ROSE/) ROSEN C A.  
 XX  
 PI Choi GH, Kunsch CA, Baraeh SC, Dillon PJ, Dougherty B, Fannon MR;  
 PI Rosen CA;  
 XX  
 DR WPI: 2002-479261/51.  
 DR N-PSDB; ABQ84899.  
 XX  
 PT New Streptococcus pneumoniae antigens, useful for detecting  
 PT Streptococcus and for preventing or attenuating disease caused by  
 PT Streptococcus infection -  
 XX  
 PS Claim 11, Page 50; 70pp; English.  
 XX  
 CC ABQ84792 to ABQ84904 represents nucleic acids which encode the  
 CC Streptococcus pneumoniae antigens given in ABP94557 to ABP94669.  
 CC The S. pneumoniae antigens have antibacterial activity and can be  
 CC used in vaccines. The S. pneumoniae antigens can also be used to  
 CC prevent or attenuate a Streptococcal infection in an animal. The  
 CC polynucleotides encoding the S. pneumoniae antigens can be used to  
 CC detect Streptococcus nucleic acids. ABQ84905 to ABQ85130 represent  
 CC primers used in the cloning of S. pneumoniae ORFs (open reading frames)  
 CC which are used in an example from the present invention.  
 CC  
 XX  
 SQ Sequence 1007 AA;  
 Query Match 94.8%; Score 4171; DB 23; Length 1007;  
 Best Local Similarity 100.0%; Pred. No. 4,4e-71;  
 Matches 627; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 38 ETSODPFKKTKRAVKEKEVSKNPVIDNNNTNEEAKIKEENSNSQGGYTDSPFNKNTEN 97  
 DB 1 ETSODPFKKTKRAVKEKEVSKNPVIDNNNTNEEAKIKEENSNSQGGYTDSPFNKNTEN 60  
 QY 98 PKKEDKVVYIAEFKDKESGEKAIKELSLKNTKVLTYTDRIENGSAIETTPDNLKIKOI 157  
 DB 61 PKKEDKVVYIAEFKDKESGEKAIKELSLKNTKVLTYTDRIENGSAIETTPDNLKIKOI 120  
 QY 158 EGISSVERBAQKVQPMNMHARKEIGVEAIDYLSINAPFGKNPDGRGMVISNIDGTIDYR 217  
 DB 121 EGISSVERBAQKVQPMNMHARKEIGVEAIDYLSINAPFGKNPDGRGMVISNIDGTIDYR 180  
 QY 218 HKARIDDDAKASRFKKEDLKGTDKNWLSDKIPHANVYNGSKITVEKDDGRDYFDP 277  
 DB 181 HKARIDDDAKASRFKKEDLKGTDKNWLSDKIPHANVYNGSKITVEKDDGRDYFDP 240  
 QY 278 HGMHIAAGLADNTEODIKNFNGIDGIAIPNAQIFSYKMYSDAGSGFAGDETFPAIEDSI 337  
 DB 241 HGMHIAAGLADNTEODIKNFNGIDGIAIPNAQIFSYKMYSDAGSGFAGDETFPAIEDSI 300  
 QY 338 KHNADVVSVSSTGFTGTGLVGEKRYQAIRALRKAGIPMVVATGNYATSSASSSMDLVANNH 397  
 DB 301 KHNADVVSVSSTGFTGTGLVGEKRYQAIRALRKAGIPMVVATGNYATSSASSSMDLVANNH 360  
 QY 398 LKMTDGTGVTAAHEDAIIVASAKNQVEPDKNIGESKYNIGACFPFKSKITTNEO 457  
 DB 361 LKMTDGTGVTAAHEDAIIVASAKNQVEPDKNIGESKYNIGACFPFKSKITTNEO 420  
 QY 458 GTKAPSKLKFVYIGKQDQDILGLDRGKIIVMDRIYKDKNAFKKAMDGARAIVNVN 517  
 DB 421 GTKAPSKLKFVYIGKQDQDILGLDRGKIIVMDRIYKDKNAFKKAMDGARAIVNVN 480  
 QY 518 TVNYNNDNMTLPAWGAEDEGTQSQVFSISGDDGVKLMNMINDPKKTEVRNNKEDFK 577  
 DB 481 TVNYNNDNMTLPAWGAEDEGTQSQVFSISGDDGVKLMNMINDPKKTEVRNNKEDFK 540  
 QY 578 DKLEQYPIIDMESFNSKNPNVGEKEIDFKAPPTDKLYKEDIIVPAGSISWGPRIIDL 637  
 DB 541 DKLEQYPIIDMESFNSKNPNVGEKEIDFKAPPTDKLYKEDIIVPAGSISWGPRIIDL 600  
 QY 638 LKPDVSAFGKNIKSTLVINGKSTGYG 664

DB 601 LKPDVSAFGKNIKSTLVINGKSTGYG 627  
 RESULT 6  
 ID AAU05396 standard; Protein; 26926 AA.  
 XX AAU05396;  
 AC AAU05396;  
 XX 24-OCT-2001 (first entry)  
 DT 24-OCT-2001 (first entry)  
 XX Human titin (connectin) protein sequence.  
 DE Human titin (connectin) protein sequence.  
 XX Human; titin; connectin; pickwick mutation; cardiac specific exon; N2B;  
 KM titin-related disease; zebrafish; heart failure; heart disease.  
 OS Homo sapiens.  
 XX WO200151666-A1.  
 PN 19-JUL-2001.  
 PD 12-JAN-2001; 2001WO-US01212.  
 XX 12-JAN-2001; 2001WO-US01212.  
 PF 12-JAN-2000; 2000US-0175787.  
 XX 12-JAN-2000; 2000US-0175787.  
 PR (GEHO ) GEN HOSPITAL CORP.  
 PA Fishman MC;  
 XX Fishman MC;  
 PI WPI: 2001-451869/48.  
 DR N-PSDB; AAS05390.  
 XX  
 DR N-PSDB; AAS05390.  
 XX  
 PT Determining if a subject has or is at risk of developing a  
 PT titin-related disease or condition, particularly heart failures,  
 PT comprises detecting the presence of a mutation in the titin gene -  
 XX  
 PS Disclosure; Page 57-111; 114pp; English.  
 XX  
 CC The present sequence representing human titin (also known as connectin)  
 CC is described in an invention relating to a novel method for determining  
 CC whether a subject has or is at risk of developing a titin-related  
 CC disease or condition. The method comprises analyzing a nucleic acid of  
 CC a sample from the subject and detecting the presence of a mutation  
 CC (e.g. the pickwick mutation in the cardiac specific exon N2B) in the  
 CC titin gene, which indicates that the subject has or is at risk of  
 CC developing a titin-related disease. The zebrafish which has a phenotype  
 CC similar to mammalian heart failure is used as a model. The method is  
 CC useful for detecting an increased likelihood of heart disease, such as  
 CC heart failure, in a patient, so that appropriate intervention can be  
 CC instituted before any symptoms occur. The method may also be used to  
 CC facilitate determination of etiology of an existing heart condition,  
 CC such as heart failure, to identify compounds that can be used to treat  
 CC or prevent heart conditions, in prenatal genetic screening, e.g. to  
 CC identify parents who may be carriers of a recessive titin mutation.  
 CC Compounds identified using the methods may be used to treat patients  
 CC that have or are at risk of developing heart disease, e.g. heart  
 CC failure.  
 CC  
 XX  
 SQ Sequence 26926 AA;  
 Query Match 70.8%; Score 3114.7; DB 22; Length 26926;  
 Best Local Similarity 9.5%; Pred. No. 4.3e-41;  
 Matches 583; Conservative 69; Mismatches 10; Indels 5463; Gaps 479;  
 QY 1 Y-----TS-S-----S---E-D-----ALN---ISD-----KEX-- 22  
 DB 10392 YCVVNSTGSRKFCQVNVVDHPGPPVGSFDEVTKQVIMVSMKPPLDGSGSKITNYI 10451  
 QY 8 -----TS-S-----S---E-D-----ALN---ISD-----KEX-- 22  
 DB 10452 IEKKEVGKDVMMPTVSASATTCVKSLTEGKDYIFRIHAEINLYGISDPLVSDSMKAKDR 10511

Qy	23	--V--A--	E--	NK--	---EK--	---H--	30
Db	10512	FRVPDADQPIVTEVTKDSALVTWKNPHDGGKPI	TNYIILEKRETM	SKRWARVTKDIHBY	10571		
Qy	31	---EN--	II	---I--	33		
Db	10572	TKFRVPDLLEGCOYEFVRVAENEIGIDPSPSKVPFAKDPIAKPSPVPNPALDITCNS	10631				
Qy	34	---H--S--	---A-M--	ET--	39		
Db	10632	VDLTWQPPRHGSGKILGIVVEYQKVGDBEWRANHTPESCPETKYKVTGLRDGQTYKFR	10691				
Qy	40	---S--	---Q--D--	---F--	43		
Db	10692	VLANAAGESDPAHVPEPVLVKDRLEPPELILDANMARQCHIKVGDITLRSALIIKGVPPP	10751				
Qy	44	---KE--K--	---T--A--V--IK--	52			
Db	10752	KVTWKKEDRDAPTAKRIDVTPVGSKLEIRNAAHEDGGIYSLTVENPAGSKTVSVKVLVLD	10811				
Qy	53	---E--	---KE--	---V--	59		
Db	10812	KPGPPRDLEVSEIRKDCYLTWKEPLDDGGSVITNVVVERRDVASAOWSPLSATSKKSH	10871				
Qy	60	---N--	---P--V--I--	---D--	64		
Db	10872	FAKHLNEGNYLFRVAENQYGRGPVETPKPIKALDPLHPPGPKDLHHVDVDTESVL	10931				
Qy	65	---N--	---N--T--S--	---N--	69		
Db	10932	VNKPDRDGGSPITGYLVYQEGTQDWIKFKVTNTLVCVVTGLQOGKTYRFRVAENIV	10991				
Qy	70	---E--	---E--	---AK--	73		
Db	10992	GLGLDPTTPIECQEKLVPPSVELDVKLIIEGLVVKAGTTRFPALIRGVVPVTAKTWTDG	11051				
Qy	74	---IK--E--	---N--	---SN--K--	81		
Db	11052	SEIKTDEHYTVETDNFSSVLTIKNCLRRDTGEYQITVSNAGSKTVAVHLTVLVDVGPPT	11111				
Qy	82	---S-Q--	---G--D--	85			
Db	11112	GPINILDVTPEHMTISWQPPKDDGSPVINIYVEKQDTRKDTWGVVSSGSSKTKLIPH	11171				
Qy	86	---YT--	---DS--	---F--	90		
Db	11172	QKGEYVFRVAENKIGVGPLDSTPTVAKHKFSPSPGPKPVVTDITENAATVSWTLPK	11231				
Qy	91	---V--	---V--	---E--	96		
Db	11232	SDGSPITGYMERREVTGKVRVKNKTPITADLKFRTVGLYEGNTYEFVRVAENLAGLSKP	11291				
Qy	97	---NPK--K--E--D--	---K--	---V--	105		
Db	11292	SPSSDPIKACRPIKPPGPINPKLKDKSRETADLWTKPLSDGSGPILGYVVECOQPGTA	11351				
Qy	106	---Y--	---I--A--	---E--F--	112		
Db	11352	QMNINKDELIRQCAFVRPGLIEGNEVFRFRIKAANIVGEGEPRELAESVIAKDILHPPEV	11411				
Qy	113	---K--E--	---S--G--	---EK--	118		
Db	11412	ELDVTCRDVITVRVGTIRILARVKGREPDITWTKEGKVLVREKRVLDLQDLPRVELOI	11471				
Qy	119	---AI--	---KE--	---L--	123		
Db	11472	KEAVRADHKYIISAKNSGHAOGSAIVNVLDRLPGPCQNLKVTNVTKENTISWENPLDN	11531				
Qy	124	---S--S--	---L--K--	---N--	129		
Db	11532	GGSEITNFIVEYRKPQKGSIVASDVTKRLIKANLLANNEYFRVCAENKVGVGPTIET	11591				
Qy	130	K--	---V--	---L--Y--	133		
Db	11592	KTPILAINPIDRPGPENLHADKGTFFYLKWRRPDYDGGSPNLHYHVERLKGSDDE	11651				
Qy	134	---T--Y--DR--	II	137			
Db	11652	RVHKGSIKETHYVMDRCVENQIYEFVRVQTKNEGGSDEWKTVEVVVVKEDLOKPVLDLKL	11711				
Qy	138	---I--	138				
Db	11712	GVLTVKAGDTIRLEAGVRGKPPPEVAWTKDADTLTRSPRKIDTRADSSKSLTKAKR	11771				
Qy	139	---F--	---N--	---G--	141		
Db	11772	SDGGKYVVTATNTAGSFVAYATVNVLDKPGPVNRNLKIVDVSSDRCTVCWDPDDGGCEI	11831				
Qy	142	---SA--	---IE--	---T--	146		
Db	11832	QNYIILEKETKRMVWSTYSATVLTPTGTTVTRLIEGNEYIFRVAENKIGTGPPTESKPVI	11891				
Qy	147	---T--PD--	---N--LD--	---K--	153		
Db	11892	AKTKYDKPGRDPPPEVTKVSKEEMTVVWNPPEYDGGKSITGYFLEKKEKHSTRWVPVNS	11951				
Qy	154	---I--K--	---Q--I--E--GI--	---S--	161		
Db	11952	AIPERMKVQNLDPDHEYQFRVKAENEIGIEPSPVVAKODIEBPPTNFRVDT	12011				
Qy	162	---S--	---VE--	---R--AQ--K--	168		
Db	12012	TGHSITLGMKPVYDGGAPIIGYVVMRPKIADASPDEGKFCNAQAOLVRKEFTVTSLD	12071				
Qy	169	---V--Q--	---P--	---M--	172		
Db	12072	ENOYEFRVCAONQVIGRPAELKEAIKPKETLEPPEIDLOASMRKLVIVRAGCPIRLPA	12131				
Qy	173	---M--	173				
Db	12132	IVRGPAPKVTRKVGIDNVVRKGOVDLVDTMFLVIPNSTRDDSGKYSLTUVNPAGEKA	12191				
Qy	174	---N--	---H--	---A--RK--	178		
Db	12192	VPVNVRLDTPGPVSLKVDVTKTCHVSWAPPENDGGSQVTHIVEKREADRKTWSTV	12251				
Qy	179	---EI--	---GV--	---E--	185		
Db	12252	TPEVKKTSFHTNLVPGNEYFRVTAVNEYGPGVTPDKPVYASDPLSEDPDRKLEAT	12311				
Qy	186	---ID--	---Y--	---LK--	190		
Db	12312	EMTKNSATLAWLPPURDGAIDGVIISYREEEQPADRWTEYSVVVKDLSLVVTGLKEGKK	12371				
Qy	191	---SI--	191				
Db	12372	YKFRVAARNAVGVSIPREAGVYEAKEQLLPPKILMPQITIKAGKKLRIEBAHYGKPH	12431				
Qy	191	---	192				
Db	12432	TCKWKKEDEVVTSSTHSLAVHKADSSILI IKDVTBKDSGYSLTAENSSGTDQIKVWV	12491				
Qy	193	---	192				
Db	12492	MDAPGPPPPDISIDADACSLSWHII PLEDGGSNITNIVEKCDVSRGDWWTALASVTK	12551				
Qy	193	---N--	---A--PFG--KN--	199			
Db	12552	TSCRVGKLIPGOEYIFRVAENRFGISBPLTSPKVAQOFFPGVPEPKNARVTKNKDCI	12611				
Qy	200	F--DG--	---R--	---GMV--	208		
Db	12612	FVAMDRPDSGGSPIIGYLIERKERNLLWVKANDTLVRSTEYPCAGLEYSFRIYA	12671				
Qy	209	---N--	---ID--	---GT--	214		

Db 12672 LNKAGSSPPSKPTXYTARMPVDPGKPEVIDVTKSTSLIWARPKDGGSKIIGYFVEA 12731  
 QY 215 -D-Y-R-----H-----KA-----M- 221  
 Db 12732 CKLBGDKRCNCRNTAHPQIPOSEYTATGLEEKAQYQFRAIARTAVNISPPSPDPVTILA 12791  
 QY 222 -RID-----D-----D-----A-----KASW-R 232  
 Db 12792 ENVEPRIDLSVAMKSLTAKAGTNCDAITVFGKMPVSMKKGITLKRPEGIKMMQOR 12851  
 QY 233 -F---KKE-DL---K---G---T---DK-----N---Y----- 245  
 Db 12852 NLCTLELFSVNRKDSGYTITAENSSSGSKSATIKLVLDKGPASVXINMGYDRAWLS 12911  
 QY 246 W--L---S-----DK-----IP-----H-AF-----N-Y 257  
 Db 12912 WEPLEDGGSEITNYIVDKRETSRPNMAQVSATVPITSVCEKLEIEGHEVQFRICAENKY 12971  
 QY 258 -----Y-----N-----GG-KIT----- 264  
 Db 12972 GVGDPVTEPAIAKPIPDPCRCPPVLSNITKDHMTVSMKPPADDGSSPITGYLEKRE 13031  
 QY 265 -V-----E----- 266  
 Db 13032 TQAVNMTKVNKKPIIERTLKATGLOEGTEYFRVTAINKAGPGKPSDASKAAYARDPQY 13091  
 QY 267 -K-YD-----DG-----R 272  
 Db 13092 PAPAPFKVDYDTRSSVSLWKGPAIDGSPPIIGYLVEVKRADSDNVRNCLPQNLQKTR 13151  
 QY 273 -D-Y-F-----DP-----H-----GMH-----I-AG 284  
 Db 13152 FEVTGLMBDYOYQFRVAVANKIGYSDPSDVPDKHYPPKDILIPRGEHDADLRKTLILRAG 13211  
 QY 285 -----I-----L-----AG----- 288  
 Db 13212 VTKMLYVPVKGRPPPKITWSKPNVNLDRIGLDIKSTDFDTFLCENNVKRYDAKYLITL 13271  
 QY 289 -N-----DT-----E-----Q-D- 294  
 Db 13272 ENSGCKKEYTIVVVLDPGPPINVTVEISKDSAYVTWEPPIIDGSPPIINYYVQKDA 13331  
 QY 295 -I-----K-----N-----F-N-GI-D-G-----I-A----- 305  
 Db 13332 ERKSWSTVTTECSKTSFRVPNLEEGSKYFFRVFAENEGIGDPGEBTRDAVASQTPGPVV 13391  
 QY 306 -----P-----N-----A-Q----- 309  
 Db 13392 DLKRVSVSKSCSIGMKKPHDGGRIIGYVVDPLTEENKMQRVMSKLSLOYSAKDLTEG 13451  
 QY 310 -----I-----F----- 311  
 Db 13452 KEYTFRVAENENEGEPSEITTVARDDVAPDLKGLPDLCLAKENSFRLKIPKIG 13511  
 QY 312 -SYKM-----Y-----SDAG----- 320  
 Db 13512 KPAPSVSMKGEDPLADTRVSVESSAVNTLLIYDQCKSDAGKYITLLKXVAGTKSGTI 13571  
 QY 321 S-----G-----F-----A-----G-----D----- 326  
 Db 13572 SIKVVGKPGIPTGPIKPEDEVTAEMTLKMAPPKDGGSEITNYILEKSDVNNKMYTCAS 13631  
 QY 327 -ETMF-----H-----A-----I-E-DS-I-KH-----N-VDV- 343  
 Db 13632 AVQKTFTRVTRLHEGMEYTFRVSABNKYVGEGKSEBIVARHPFDVADAPPNNIVDR 13691  
 QY 344 -VSU-----S-----SG-----F-TG-T-GL----- 355  
 Db 13692 HDVSVLWTDBPKTKGTGPIYHLEFKERNSLMLKMRANKTPDIRMDPFVTLTGLELEYEF 13751  
 QY 356 -VG-----E-KY----- 360

Db 13752 RVMAINLAVGKPSLPSPEPVVALDIPDPGKPEVINITRNSVTLIMTEPKYDGHKLTV 13811  
 QY 361 -W-----Q-A-----I-----PALR-K-AG-I--P-----MV 375  
 Db 13812 IVERKDLPSKSMKANHNVPCEAFVTLVLEGGKYEF--IRAKNTAGAISAPSESTETI 13870  
 QY 376 V-----A-----T-----G-----N----- 380  
 Db 13871 ICDVEYAPPTIVLDPITIKDGLTIKAGDTYLVNAISILGKPLPKSSWSKAGKDRPSDITQ 13930  
 QY 381 -YAT----- 385  
 Db 13931 ITSTPTSMLTIKYATRKDAGEYTTATNPFGTKEHVKTVLDPGPBPVSEISNVAE 13990  
 QY 386 -S-----S-----SS-----W-----DL-----VA-----N----- 395  
 Db 13991 KATLTWTPLEDGSPKSYILEKRETSRLMTVSEDIOSCRHVAATKLIOGMEYFRVS 14050  
 QY 396 -NHL-----KMTD-----TG-----NVR-TAA-----H 412  
 Db 14051 AVNHYGGEVQSEBPVNMVDRFPGPPEKPEVSNVTNTATATYSWKRPPVDDGSEITGYH 14110  
 QY 413 -E-----D-----A-I-----A-V----- 418  
 Db 14111 VERREKSLRMVRAIKTPVSDLRCKVTGLOGSTYEFRVSANRAGICGPSEASDSVLUMK 14170  
 QY 419 -A-----SA-----K-----N-----Q-----T----- 425  
 Db 14171 DAAVPPGPPSNPHVTDTTKKSASLAWGKPHYDGLGITGVVHEKQVDEAMIKDTGTGA 14230  
 QY 426 -V-EF-----D-----K-----VN-----I----- 433  
 Db 14231 LRITQFVVPDLQTEKYNFRISAINDAGCEPAVIVPELVERBMADEFELDAELRRTLV 14290  
 QY 434 -G-----G-----ESF-----KY----- 440  
 Db 14291 VRAGLSIRIFVPIKGRAPAVETWKONINLKNRANIENTESFTLLIPECGRVOTGKFMV 14350  
 QY 441 -----R-----N-----I-G----- 444  
 Db 14351 TIENPAGKSGPVNVVRVLDPGPVNLNRPDITKDSVTLHMDLPLIDGSRITNYIVEKR 14410  
 QY 445 -A-----A-----FF----- 447  
 Db 14411 EATRKSYSTATTKCHKCTYVATGLSECEYFFRVMAENEYIGEPPTTEBPVKAESAPSP 14470  
 QY 448 -----D-----K-----SK-----ITT----- 454  
 Db 14471 PDSLINIMDITKSTVSLAMPKPKDGGSKIIGYVIEAQRKSGSDOWHTITTYKGLCEVVRNL 14530  
 QY 455 -----N-----E-D-G-----TKA----- 461  
 Db 14531 TEGEYTFQVMAVANSAGRSAPRESRPVIVKEQTMPLPELDLRIYQKLVIAKAGDNIIVEI 14590  
 QY 462 -PS-----K-----LK-----F-----Y-----I-GK- 472  
 Db 14591 PVLGRPKPTVWKKGDOILKQTORVNEFTATSTILINECVRSDSGPPPLTARNIVGEV 14650  
 QY 473 G-----Q-----D-----Q-D----- 477  
 Db 14651 GDVITIOVNDIPGPPIPIKFDEVSDFVTFMSDPENDGCVPISNVVENRQJDSITTW 14710  
 QY 478 -L-----I-----GL----- 481  
 Db 14711 ELATTVIRTTYKATRLTTLTGLEYQFRVAKQNRGVGPGITSAMIVANYPFKVPGBPPTQV 14770  
 QY 482 -----D-----L-----R-----G-----K-----I- 487  
 Db 14771 TAVTKDSMTISMEHPLSDGSPILCYHVERKERNGILMOTVSKALVPGNIFKSSGLTDGI 14830  
 QY 488 A--V-----M-----D-----R--I-----YT----- 495  
 Db 14831 AYEFRRVIAENMAGSKSKSPSEPLALDPIIDPGKPPVPLNITRHTVTLKMAKPEYTOGFK 14890

```
QY 496 -----K-DL-----KVA-----501
Db 14891 ITSIVIEKRDLPNGRWLKNFNSILENFTVSGLTEDAAYEFVRVIAKNAAGAISSPPSPS 14950
QY 502 -----PK-----KA-----M-----506
Db 14951 DAITCRDDVEAPKIKVDVKFDTVILKAGEAFRLADVSGRPPPTMEWSKDGKELEGTA 15010
QY 507 -----D-----507
Db 15011 LEIKIADFSTNLVANKSTRRDSGAYTLTATNPGGFAKHI FNVKVLDRPGPEGPLAVTEV 15070
QY 508 -----K-----GA-----R-----A-----I- 513
Db 15071 TSEKCVLSWFPPLDDGGAKIDHIYIVQKRETSRLAWTNVASEVQVTKLVTKLLKGNEYIF 15130
QY 514 -----M-----V-----VN-----T-----V 519
Db 15131 RVMAVNVKGVGEPLSEPEVLAVNPYGPDPDPKNPEVTTITKDSMVVCGHPDSDGSGSEII 15190
QY 520 NY-----Y 522
Db 15191 NYIIVERDKAGQRIWKCNKKTLDLRYKVSGLTEGHEVEYEFRIABNAAGISAPSPSPFY 15250
QY 523 -----N-R-D-----N-----W-----527
Db 15251 KACDTVFKPGPGNPRVLDTSRSSISIAWNKPIYDGGSEITGYMVEIALPEEDEMQIVTP 15310
QY 528 -----TE-----L-PA-----M-----533
Db 15311 PAGLKATSYTTIGLTENQEQYKIRIYAMNSEGLGEPALVPGTPKPADRMPLPEIELDADLR 15370
QY 534 -----534
Db 15371 KVTIRACCTLRLFVPIKGRDPPEVKWARDHGESLDKASIESASSYTLIIIGNVNRFD 15430
QY 535 -Y-----A-----D-----E-----G-----540
Db 15431 KYILTVENSNGSGAFVNRVLDTGPPQDLKVAKEVTKTSVLTWDPPLDGGSKIRNYI 15490
QY 541 -----T-K-S-----Q-----VF-----SI-S- 549
Db 15491 VEKRESTRKAYSTVATNCHTKTSWKVDQLQEGCSYFRVLAENEYIGILPAETAESVKASE 15550
QY 550 -----G-----D-----VKL-----W-----557
Db 15551 RPLPPGKITLMDVTRNSVLSWEKPEHDDGSGRIILGYIVEMQTKGSDKWATCATVKVTEAT 15610
QY 558 -----N-----M-I-----N-----561
Db 15611 ITGLIQEGEYSFRVSAQNEKGISDPRLQSVPIAKDLVIPPAPKLLFNTFTVLAGEDLKV 15670
QY 562 -----P-----D-K-----K-TE-----VK-----R-----N--- 571
Db 15671 DVPPIGRPTPAVTHKONVPLKQTRVNAESTENNSLLTIKDACREDVGHVYVVKLTNSAG 15730
QY 572 -----N-----K-E-----D-----F-----K-D-----578
Db 15731 EAIETLNVILDKPGPPTPGVKMDVNTADTSITLSWGPPEKYDGGSSINNYIIVEKRDTSITT 15790
QY 579 -----K-L-----E-Q-----Y-----Y-----585
Db 15791 WQIVSATVARTTIKACRLKTCEYQFRIAAENRYGKSTYLNSEPTVAQYPPFKVPGPPGPT 15850
QY 586 -I-----D-ME-----S-----P-----NS-----NK-----P-----596
Db 15851 VVTLSSRDSMEVQWNEPISDGGSRVIGYHLERKERNISILWKLNKTPIPTQTKFKTTGLEE 15910
QY 597 -----N-VG-----D-----600
Db 15911 GVEYEFVSAENIVIGICKPSKVSCEYVARDPCDPPRPEAIIVTRNSVTLQWKFTYDGG 15970
```

---

```
QY 601 -----EK-----E-----ID-F-----KF-----A-----P- 610
Db 15971 SKITGYIIVEKKELPEGRWMAKSFTHIIDTHFEVTGLVEDHRYEFVRVIAAAGVFSEPS 16030
QY 611 -----DT-----D-----K-----EL--- 616
Db 16031 STGALTARDEVDPPRISMDPKYKDIIVVHAGESFKVDADIYKGPPTIQWIKQDQELSNT 16090
QY 617 -----Y-----K-----E-----D-----I- 622
Db 16091 ARLEIKSTDFATSLSVKDAVRVDSGNYILKAKNVAGERSVTNVVKVLD RPPGPGPWIS 16150
QY 623 -V-----P---AGS-----TS---W----- 630
Db 16151 GVTAEKCTLAWKPPLODGGSDIINYVERRETSRLVTVVDANVOTLSCKVTKLLEGNEY 16210
QY 631 -----G-----P----- 632
Db 16211 TFRIMAVNVKGVGEPLSEPEPVAKNPFVVPDAPKAPVTTVTTKDSMIVVWERPASDGGSE 16270
QY 633 -----R-----I---DL-----L----- 637
Db 16271 ILGYVLEKRDKEGIRWTRCHKRLIGELRLRVGTGLIENHDYEFRVSAENAAAGLSEPPSPA 16330
QY 638 -----LKP-----D---V---SAP-----G---K---N----- 648
Db 16331 YQKACDPIYKPGPPNPKVIDITRSVFLSWSKPIYDGCETOGYIIVEKCDNVNGEWTMC 16390
QY 649 -----I-KST-----L-----N----- 654
Db 16391 TPPTGINK-TNIEVEKLEKHEYNFRICAINKAGVGEHADVPGPPIIVEKLEAPDIDL 16449
QY 655 -----VIN-----GK-----S-T-----Y 662
Db 16450 ELRKIINIRAGGSLRLFVPIKGRPTPEVKWKGVDGEIRDAIIDVTSPTSLSLVDNVR 16509
QY 663 --G-Y 664
Db 16510 DSGKY 16514
```

RESULT 7  
ABG74786  
ID ABG74786 standard; Protein; 31267 AA.  
XX  
AC ABG74786;  
DT 05-JUN-2003 (first entry)  
XX  
DE Human RGS11 protein.  
XX  
KW RGS11; human; screening; cardiact; antianginal; gene therapy;  
KW heart disorder; cardiac ischaemia; heart failure; angina.  
OS Homo sapiens.  
XX  
PN WO2002103355-A1.  
XX  
PD 27-DEC-2002.  
XX  
PF 17-JUN-2002; 2002WO-JP06019.  
XX  
PR 18-JUN-2001; 2001JP-0183038.  
XX  
PA (TAKE ) TAKEDA CHEM IND LTD.  
XX  
PI Koyama N, Tanida S, Yamamoto K;  
XX  
DR WPI; 2003-167557/16.  
XX  
DR N-PSDB; ABX13540.  
PT Screening compounds regulating RGS11 expression and activity for  
prevention and treatment of heart disease -

XX Claim 1; Page 59-261; 321pp; Japanese.

CC This invention describes a novel method for screening compounds for their  
 CC ability to regulate the activity and expression of human RGS11 and its  
 CC partial peptides and salts, by observing the expression or activity of  
 CC RGS11 in the presence or absence of the test compound. The products of  
 CC the invention have cardiant and antilanginal activity and can be used for  
 CC gene therapy. The methods and compositions are useful in the prevention,  
 CC treatment and diagnosis of heart disorders such as cardiac ischaemia,  
 CC heart failure and angina. This sequence represents the human RGS11  
 CC protein described in the disclosure of the invention.

XX Sequence 31267 AA;

Query Match 70.6%; Score 3104.2; DB 24; Length 31267;

Beet Local Similarity 9.4%; Pred.No.1.6e-40; Indels 5628; Gaps 490;

Matches 594; Conservative 59; Mismatches 11; Indels 5628; Gaps 490;

QY 1 Y-----PV-----V-----LAD-----7  
 Db 15617 YCVVENVSTGRKRCQCVNVVDHRCPRVGVSPFDEVTKDYNVISMKPLDDGGSKITNYI 15676  
 QY 8 -----TS-S-----S--E--D-----ALN--ISD-----KEK--22  
 Db 15677 IEKKEVGKDVMPVTSASAKTTCKVSKLEGKDYIFRIHAENLVGISDPLVSDSMKAKDR 15736  
 QY 23 --V--A-----E-----NK-----EK-----H--30  
 Db 15737 FRVPDAPDPIVTEVTKDSALVTWKNKPHDGGKPIITNYILEKRETMKSKMAVTKDPIHPY 15796  
 QY 31 -----EN-----I-----33  
 Db 15797 TKFRVPDLLECQYEFVSAENIGIGDPSPSKVFPAKDIAPSPVNEALDITGNS 15856  
 QY 34 -----H--S-----A-M-----ET-----39  
 Db 15857 VDLTPQPRHDGSKILGIVIEYQKVGEWBRANHTPESCFETKYKVTGLRDGQTYKFR 15916  
 QY 40 -----S-----Q-----D-----F--43  
 Db 15917 VLAANAAGESDPANHPPEVLVKRLPELILDANMARQHIKVGDTLRLSAIKGVFP 15976  
 QY 44 -----KE-----K-----T--A--V-IK-----52  
 Db 15977 KVTWKEDRDAPTKARIDVTVEGSKLEIRNAHEDGIGYSLTENPAGSKTVSVKVLVD 16036  
 QY 53 -----E-----KE-----VV-----SK-----59  
 Db 16037 KRGPRLDVESEIRKDSCTLTWKEPLDDGSGVITNVYERRDVASQWSPISATSKKSH 16096  
 QY 60 -----N-----P-V-----I-----D-----64  
 Db 16097 FAKHLNEGNYQLFRVAENQYGRGPFVETPPKIKALDPLHPGPKDLHHVDVKTVEL 16156  
 QY 65 --N-----N-----T-----S-----N--69  
 Db 16157 VMNRPDRDGGSPITGYLYVEDEGTQDWIKKFTVTNLECVVGLQOGKTYFRVKAENIV 16216  
 QY 70 -----E-----E-----AK-----73  
 Db 16217 GLGLPDTTPIIECEKLVPSVELDVKLEGLVYVAGTTPVFPRAIIRGVPTAKWTTDG 16276  
 QY 74 --IK--E--E-----N-----SN--K-----81  
 Db 16277 SEIKTDEHYTETDNFSSVLTIKNCRLRDTGEGYQITVSNAGSKTVAVHLTVLDPGEP 16336  
 QY 82 -----S-Q-----G-----D-----85  
 Db 16337 GPINILDVTPHEMTISWQPPKDDGSPVINYIVKQDTRKDTWGVSSGSSKTKLKIPHL 16396  
 QY 86 --YT-----DS-----F-----90

Db 16397 QKCEYVFRVAENKIGVPELDDSTPTVANKHSPSPSPCKPVVTDITENAAVSWTLPK 16456  
 QY 91 -----VNK-----NT-----E-----96  
 Db 16457 SDGSPITGYMERREVTGKMRVNRKTPPIADLKRVGLYEGNTYFRVFAENLAGLSKP 16516  
 QY 97 -----NPK--K-E--D--K-----VV-----105  
 Db 16517 SPSSDPIKACRPKIPKPPPINPKLKDSDRETADLVMTKPLSDGSPILGVVVECOKPGTA 16576  
 QY 106 -Y-----I--A-----E--F-----KD-----112  
 Db 16577 QMNRINKDELIRQCAFVPGLEGENEYFRKIANIVGEGSPRELAESVIKDTILHPPEV 16636  
 QY 113 -----K--E-----S--G-----EK-----118  
 Db 16637 ELVDTCRDVITVRVGOTIRILARVKGPEPDITWTKEGKVLVRKRVLDIODLPRVELQI 16696  
 QY 119 -----AI-----KE-----L--123  
 Db 16697 KEAVRADHGKTIISAKNSSGHAQGSALVNVLDNRGPCQNLKVTNTKENTISWENPLDN 16756  
 QY 124 -----S--S-----L-K-----N-----T--129  
 Db 16757 GGEITNFIYEVRKPKNGKWSIVASDVTKRLIKANLLANNEYFRVCAENKVGVPITET 16816  
 QY 130 K-----V-----L-Y-----133  
 Db 16817 KTPILAINPIDRGPENLHIDKGTFFVYKWRPRPDYDGGSPNLSTYHVERRLKGSDDWE 16876  
 QY 134 -----T-Y--DR-----I--F-----N--G-----S--142  
 Db 16877 RVHKGSIKETHYMDRCEVNGIYEFRTQTKNEGSESDMVKTEBVVVKEDLQKPLVLDKLS 16936  
 QY 143 -----A--I-----E-----T--T--P-----148  
 Db 16937 GVLTVKAGDTIRLEAGVGRKPFPEVAMTKDADTLTRSPVKIDTRADSSKFSLTRAKR 16996  
 QY 149 -D-----N-LD-----KI-----K--155  
 Db 16997 SDGKYVVTATNTAGSFVAAYATVNVLDKPGFVRNLKIVDSSDRCTVCMWPPEDDGCET 17056  
 QY 156 Q-----IG-----I--160  
 Db 17057 QNTILEKCETKRMWSTYSATVLTPTGTVTRLLIEGNEYIRVRAENKIGTGPTESKPVI 17116  
 QY 161 -----S-----SV-----E--R-----165  
 Db 17117 AKTKYDKRPRDPPEVTKVSKEBVTVMNPPEYDGSINGYFLEKKEKISTRAVVPVKS 17176  
 QY 166 A---Q-KVQ-----P-----171  
 Db 17177 AIPERRMKVQNLPRDHEYOFKVAENIGIGESLPSRPVAVKADIEPPGPTFRVVD 17236  
 QY 172 -----M-----N-----H-----A-RK-----178  
 Db 17237 TKGISITLGWCKPVYDGGAPIIIGYVEMRPKIADASPDEGCKRCAAAQLVRKEFTVSLD 17296  
 QY 173 -----M-----N-----H-----A-RK-----178  
 Db 17297 ENGEYFRMGLKHHGHVSWAPREDDGSQVTHYIVKRBADRKWTSTVTPVKKTSFHV 17356  
 QY 179 -----E-----E-----179  
 Db 17357 NLVPGNEYFRVAVNEYGPGVPTDVPKPVLASDPLSEPPPKLVEVTEMTKNSATLAWL 17416  
 QY 180 -----I-----E-----180  
 Db 17417 PPLBDGAKKIDGYITTSREEQPADRWTEYSVKDLSLVTTGLKEGKKYKFRVAARNAV 17476  
 QY 181 -----GV-----E-----EA-----185  
 Db 17477 VSLPRAEGVYEAKEQLLPPKILMPEQITIKAGKULRIEAVVYVGPKPHPTCKMKKGDEBV 17536

QY	186	-----	185
Db	17537	TSSHLAVHKADSSSILIIKDVTROSGYYSLTAENSSGTDQKIKVVMNDAPGPPQPPD	17596
QY	186	ID-----Y-----L-----K-S-----	191
Db	17597	ISDIDADACSLSWHIPELGGSNITNYIIVEKCDVSRGDWVTALASVTKTSCRVGKLI	17656
QY	192	I-----N-----A-PFG-----KN-----F-----DG-	202
Db	17657	EVIFRVAENRRGISEPLTSPKMAQOPRGVSEPKNARVTKNKDCIFVAMDRPDSGG	17716
QY	203	-----R-----GMV-----I-S-N-----	209
Db	17717	SPIIGYLIBERKERNLLWKANDTLVRLTEYPCAGLVEGLEYFRYIALNKAGSPSPK	17776
QY	210	ID-----T-----GT-----D-Y-R--	217
Db	17777	TEYVTARMPDPGKPEVIDVTKSTVSLIWARPKHDGSKIIGYFVEACKLPDCKWVRN	17836
QY	218	H-----KA-----M-----RID-----	224
Db	17837	TAPHOIPOEYATATGLEEKAQYQFRAIARTAVNISPPSEPSDPVTILAENVPPRIDLSVA	17896
QY	225	-----D-----A-----KSM-R-----F---K	234
Db	17897	MKSLLTVKAGTNVCLDATVFGKPMPTVSWKDGTLKPAEGIKMAQORNLCLELFSVRN	17956
QY	235	KE-DL--K-G--T--DK--N-Y--W--L--S--	248
Db	17957	KOSGDYTTAENSSGKSATIKLVLDKGPASVKINKYSDRAMLSWEPPLEDGSGEI	18016
QY	249	DK-----IP-----H-AF-----N-Y-----	257
Db	18017	TNVIDKRETSRPNQAQVATPITSCSVKEKIEGHEYQFICAENKYGVDPVFTEPAI	18076
QY	258	Y-----N-----GG-KIT-----V---	265
Db	18077	AKNPYDPPGRCPPVISNITKDHMTVSWKPPADGGSPITGYLLEKRETOAVNWKVNRK	18136
QY	266	E-----E-----K-YD- 269	
Db	18137	PIIERTLKATGLOEYTEYFRVTAINKAGPGKPSDASKAAYARDPOYPPGPAFPKYDT	18196
QY	270	DG-----R-D-----Y 274	
Db	18197	TRSSVSLSGKPAYDGGSPIIGYLVVEVKRADSDNMVRNCLPNLQKTRPEVTGLMEDTQY	18256
QY	275	F-----DP-----H-----G-M-----	280
Db	18257	QPRVAVNKIGYSDPSDPDKHYPKDILIPPEGELDADLRKTLILRAGVTMLRYVPVKGR	18316
QY	281	H-----I-----AG-IL-----	286
Db	18317	PPPKITWSPKNVNLRRDRIGLDKSTDFDPLRCENVKNKYDAGKYILTLNENSGCKEYTV	18376
QY	287	A-----G-----N-----D-----TB- 292	
Db	18377	VKVLDTPGPPVNVVTKIISKDSAYVTWEPPIIDGGSPINIVVYVQKRDABKSKSWTVTTEC	18436
QY	293	-----Q-----D-----	294
Db	18437	SKTSFRVNLBEGKSYFRVFAENYIGIDPGGETRDAVKASQTPGPPVVDLKVRSVSKSSC	18496
QY	295	I--K-----N-----F-----	298
Db	18497	STGWKAPHSDGSGSRIIGYVVDLFTENKQVRNKSLSLOYSAKDLTEGKEYTFRVSAENE	18556
QY	299	NG-----I-----D-G-----I-----AP-----	306
Db	18557	NGEGTSEITVWARDVAVPDLDLKGLPDLCLYLAKENSFRLKIPIKGKPAPSVSMKGE	18616
QY	307	NA-----Q-----	309
Db	18617	DPLATDTRVSVESSAVNTLLIVYDCKQSDACKYTTILKNVAGTKEGTISIKVVGKPGIPT	18676
QY	310	I-F-----S-----Y-----K-----	314
Db	18677	GPIKDEVTAEBATLKWAPPKDDGSEITNYILEKRDVNNKWTCASAVQKTTFRVTRL	18736
QY	315	M-----Y-----S-----DA-----	319
Db	18737	HEGMEYTFRVAENKYGVGEGLKSEPIVARHPFDVPDAPPENIVDVHRHDSVSLTWDPK	18796
QY	320	GS--G--F-----AG-- 325	
Db	18797	KTGSGPITGYHLEFERNKSLWKRANKTPIRMDFKVTGLTEGLEBYFRVMAINLAVGK	18856
QY	326	D-E-----	327
Db	18857	PSLPSEPVVALDPIEAPTIVLDPTIKDGLTIKAGDTIVLNAISILGKPLPKSSMSKAGKD	18916
QY	328	T-M-----F-----HA-----	332
Db	18917	IRPSDITQITSTPTSSMLTIKYATRKDAGEVITATNPFGTKVEHVKVTVLDVPVPGPV	18976
QY	333	I-----ED-S-IK-----H-----	339
Db	18977	EISNVSAEKATLTWTPPLEDGGSPISKYILEKRETSRLLTWTVSEDIOSCRHVATKLIQ	19036
QY	340	N-----VD-----V-----VS-----	345
Db	19037	NEYIFRVAENVHYKGEPVQSEPVQVDRFGPPGPEKPEVSNVTNTATVSKRPVDDG	19096
QY	346	V-----S-----G-----F-----	350
Db	19097	GSEITGYHVERREKSLRWRAIKTPVSDLRCKVTGLOEGSTYEFVRVAENRAGIGPPNP	19156
QY	351	T-----G--TGLV-----G-----	357
Db	19157	PGPPNPVHTDTTKSASLANGKPHYDGGLEITGYVVEHOKVGDAMIKDTGTALRITQ	19216
QY	358	EKY-----	360
Db	19217	FVVPDLQTKENYFRISAINDAGVGEPAVDPDVEIVEREMAPDPFELDAELRRLVVRAGL	19276
QY	361	W-----QA-----I--R-----	365
Db	19277	SIRIFVPIKGRPAPEVTWTKONILKNRANIENTESFTLLIIPECNRYDTGKFVMTIENP	19336
QY	366	A-----LR-----	368
Db	19337	AGKKSGFVNVRLDTPGPNLRLPDTITKDSVTLHWDLPDIDGSGRITNYIIVEKREATR	19396
QY	369	K-----A--GI-----P-----	373
Db	19397	SYSTATTKCHKTKYKVTGLSECEYFFRVAENYEGIGETPETTEFPVKAASEAPSPDSLN	19456
QY	374	M-----VV-A-----T-G-----N-----	380
Db	19457	IMDITKSTVSLAWPKPKHDGSKITGYVIEAQKSGSDQWTHITTVKGLECVVRNLTEGEE	19516
QY	381	Y-----A-TSA--S--S-----	387
Db	19517	YTFQMVNVSAGRSAPRESRPVIVKEQTMPLPELDLRIYQKLVIAKAGDNKIVEPVLGR	19576
QY	388	S-SW--D--L--V--A-----N-----	396
Db	19577	PKPTVTKKGOILKQTORVANFETATSTILNINECVRSDSGPYPLTARNIVEYGDVIT	19636
QY	397	H-----LK-----M--TD-----	402
Db	19637	IQVHDIQGPPTGPIKDFEVSDFVTFSWDPPENDGGVPISNVVEMRQDSTTWVELATT	19696
QY	403	-----TG-----N-----V-----T-----	407

```

Db 19697 VIRTAKATRLTGTLEQFVRKAGNRVGVGITSACIIVANVPFKVPGPQIPQIRPYR 19756
Qy 408 -----RT-----A-----A-----H-----412
Db 19757 ARIAEFVRVAENNAAGSKSPKSPBMLALDIPDGPVPLNTRHTVTLKMAKPEYTG 19816
Qy 413 -----E--D-----A-----I-----AV-----AS-----AKN-----423
Db 19817 GFKITSYIVEKRDLPNGRWMLKANFSNLINENEFTVSGLTEDAAIEPRVIAKAAAGAISSPS 19876
Qy 424 -----Q-----Q-----TV-----E--F-----D-----429
Db 19877 EPSDAITCRDDVEAPKJIKVDVKFKDVTILKAGAEAFRLBADVSGRPPTMEWSXOGKELEG 19936
Qy 430 -----K-----VN-----I-----I-----433
Db 19937 TAKLEIKIADFTNLVNMKDSRRDSGAYTLTATNPGFPAKHIFNVKVLDRPPEGPLAV 19996
Qy 434 -----GG-----E-----S-----437
Db 19997 TEVTSEKCVLSWFPPLDGGAKIDHYIVOKRETSRLANTVASEQVTKLVTKLKONE 20056
Qy 438 -----KX-----KX-----440
Db 20057 YIFRMAVNAKYGVGEPLESEPVLANPYGPPDPKNPEVTITKDSMVVCMGHPDSDGGS 20116
Qy 441 -----R-----N-----I-----G--A-----445
Db 20117 EIIIVYERRDKAGORWIKCKNKKTLTDLRYKVSGLTEGHEVEFRIMENAGISAPSPTS 20176
Qy 446 -----F-----D-----K-----S-----450
Db 20177 PFYACDTVFRRPGPNRVLDTSRSSISIAMNKPIYDGSSEITGYWEIALPEDEKQI 20236
Qy 451 -----K-----IT--T-----N--E--D-----GT--KA-----PS--KL--465
Db 20237 VTPRAGLAKATSYITLTGLENQYKIRIYANNSGEGEPALVGPTRKADRMALPEIEBDA 20296
Qy 466 -----K-----FV-----Y-----I-----470
Db 20297 DLRKVVITIRACCTRLRLFVPIKGRPAPEVMKARDHGESLDSKASIBSTSSYILLIGNVNR 20356
Qy 471 -----GK-----G--D-----Q--D-----ODL-----I-----G-----480
Db 20357 DSGKYLITVENSSGSKSAFVNVRLDTPGPQDLKVKETVTSYTLTWDPPLDGGSKIK 20416
Qy 481 -----L-----L-----D-----L-----483
Db 20417 NIYIEKRESTRKAYSTVATNCHKTSWKVDQLOEGSSYFVRVLAENEYIGILPAETAESVK 20476
Qy 484 -----R-----GKIAVMD-----RI--Y-----494
Db 20477 ASERPLPFGKITLMDVTRNSVSLSWEKPEHDGSRILGIYIVEMQTGSKDMKATCATVKT 20536
Qy 495 -----T-----T-----KDL-----K--N-----A-----501
Db 20537 EATITGLIOGEEYFSRVAQNEKGISDPRQLSVPIALDLVPPAFKLLFTFTVLAGE 20596
Qy 502 -----F-----F-----K-----K-----504
Db 20597 LKVDVPGFGRPTPAVTHKDVNPLKQTRVNAESTENNSLLTIADACREDVGHVVKLTN 20656
Qy 505 -----A-----A-----MD-----K--G-----509
Db 20657 SAGEALETIANVILDKPGPTPGVMKDEVTADSLTWSGPKYDGGSSINNVIVEKRDTS 20716
Qy 510 -----AR-----A-----I-----513
Db 20717 TTTMOIVSATVARTTIKACRLKTGCEYQFRIAENRYSKSTYLNSEPTVAOYPPKVPGP 20776
Qy 514 -----M-----M-----V-----NT--V--N-----520

```

```

Db 20777 GTPVVTLSRDSMEVQWNEBISDGGSRVIGYHLECKERNISLWKLNKTPIPQTKFXTTG 20836
Qy 521 -----Y--Y-----N-----RD-----N-----W-----527
Db 20837 LEBGEVEYFRVSAENIVIGICKPSKVSCEYVARDCDPGRPEALIVTRNSVTLQMKPTY 20896
Qy 528 -----T-----ELPA--M-----G-----YE-----536
Db 20897 DGSKITGYIVEKKELEBEGMKASFTNIIDTHEVGLVEDHRYEPRVIAARNAGVFSE 20956
Qy 537 -----A--DE-----G-----T-----K--SQ--544
Db 20957 PSESTGAITARDEVDPRISMDPKYKDTIVVHAGESFKVDADYKXP1PTIOWIKGOEL 21016
Qy 545 -----V-----F--SI-----SG-----D-----551
Db 21017 SNTARLEIKSTDFATSLSVDAVARVDGNYILKAKNAGRSYTVNVKVLDRPPEGPV 21076
Qy 552 -----DG-----VK-----L--W-----557
Db 21077 VISGVTAECTLAMKPEPLDGGSDIINYIVERETSRLVTVVDANQVTLSCVTKLLEG 21136
Qy 558 N-----M--IN-----PD--K-----K-----565
Db 21137 NEYFRIMAVNAKYGVGEPLESEPVVAKNPFVDPAPAPAEVTVTKDSMIWVMBRPA5DG 21196
Qy 566 --TE--V-----KR-----N-----N--KE--574
Db 21197 GSEILGYLEKRDKEGIRWTRCHRLIGELRLVTLGIEHNDYEFVSAENAGLSESP 21256
Qy 575 -----D-----F--K--D--KLE--Q--Y-----583
Db 21257 PSAYQACDPIYKRPNNRVIDITRSSVFLSMKSKPIYDGGCEIOGYIVEKCVSVGEW 21316
Qy 584 -----Y-----PI-----586
Db 21317 TMCPTPTGINKNTIEVEKLEKEHEYNFRICAINKAGVGHADVGP1IVEKLEAPDIDL 21376
Qy 587 DME-----SF-----N--592
Db 21377 DLELRKIIIRAGSLRLFVPIKGRPTPEVKMGKVDGEIRDAIIDVTSSTSLVDNVN 21436
Qy 593 -----S-----N--K-----P--N-----V--G--D-----600
Db 21437 RYDSGKTYLLENSSGTSKSAFYTVRVLDTPSPVNLKVTETIKDSVSITWEPPLDGGSK 21496
Qy 601 -----EK--E-----ID-----F-----606
Db 21497 IKNYIVEKREATRKSVAAVVNTNCHKNSWKIDQLOEGSSYFVRVTAENEYIGILPAQTADP 21556
Qy 607 --KFA--P-----D-----T-----D-----K--E-----615
Db 21557 IKVAEVPQPGKITVDVTRNSVSLSWTKPEHDGSKIIYIVEMQAKHSEKSECARVK 21616
Qy 616 -----L-----Y-----KE--D-----I-----I--V--PA-----G--626
Db 21617 SLOAVITNLTOGEERYLERVVAVNEKGRSDPRSLAVIIVAKDLYIEPDKPAFFSSYVOVG 21676
Qy 627 -----S-----T--SW--G--P-----RI-----DL--L-----637
Db 21677 QDLKIEVPISGRPKPTITWTKDGLPLKQTRRINTVDSLDLTTLSIKETHKDDGGQYGITV 21736
Qy 638 -----L--KP-----DVSA-----P--G--KN--I--K--650
Db 21737 ANNVGOKTASIEIWTLPKPPKPGPVKFDVSAESITLSNPNPLYTGCGCOITNYIVOKRD 21796
Qy 651 -----S-----TL-----N-----VI-----656
Db 21797 TTTTMDVVSATVARTTLKVTKLKTGTEYQFRIAENRYSKSFALBESDPIVAOYPRKEPG 21856
Qy 657 -----N-----N-----G-----S--T--Y-----GY--664
Db 21857 PRGKASNSSECYVARDCDPGPTEPIWVKNBETLQWTKRPVVDGSGMIGY 21908

```

RESULT 8

ABB58144  
ID ABB58144 standard; Protein; 7107 AA.

XX AC ABB58144;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 1224.

XX KW Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE ) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX DR N-PSDB; ABL02247.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

XX PS Disclosure; SEQ ID NO 1224; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).

XX CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 7107 AA;

Query Match 70.4%; Score 3098.2; DB 22; Length 7107;  
Best Local Similarity 9.7%; Pred.NO. 6.1e-44;  
Matches 579; Conservative 74; Mismatches 10; Indels 5278; Gaps 475;

QY 1 Y-----P-VV-----L--A-----D-----7  
DB 1020 YQFRVIAVNAKGLSPSPDASVPQIVKYKLLKPRIDRSNLKPLIRAGPIRYDVNVNGEP 1079  
QY 8 ----T-----SSSE-----D-A- 14  
DB 1080 APVITWYQNDKELKPEELPSSEIKNIPYNTKISIIETVRKHTGIYKIIAVNEHQDEAT 1139  
QY 15 --LNI-----S-----D-----K-E-----K-V-A---E-----25  
DB 1140 VEVNIIAPPSPKRGPLDVKDVTKDSCKLKWKKPEDDGGKPIAYQVEKFDKQGRWVPLG 1199  
QY 26 ----N-----K---EKHE-----NI- 33  
DB 1200 RTSANDTEFDVKLGQEGHEYQFRVKAINEGESDPLSDSDSIIAKNPYDAASKPCTPNIV 1259

QY 34 ----H-----S-A-----M-----E-----TS-----40  
DB 1260 DYNEHWVKLWKEAPRSOGAPISGYIIEKKOKFSPIWDEILSTNTSVPEATVEGLVEGNI 1319  
QY 41 -Q-----D-----F-----K-E- 45  
DB 1320 YQFRVRAVNAKAGSPSPDSEPHLAKPRNLKPYINRDKWKPIKVRAGQPVKFDVVKGP 1379  
QY 46 ----K-----K-----T-----A-----49  
DB 1380 APSLTWFLKETELTSTGOVRLNIDYNTKLTLLDTRKQSGQYKLAENINGVDEAVVEV 1439  
QY 50 ----K-----K-----VI-----K-----52  
DB 1440 IILDKPSKPEGPIEVSDIHKEGCKLWKRKPKDDGIGIPITGYVIEKMDTATGKWVAGSVD 1499  
QY 53 -EK--E-----V--V-----S-----KNP-----61  
DB 1500 PEKYDIEIKGLDPNHRVQFRVKAIVNEEGESEPLETESAITAKNPFDFVSAPPGLEPELWD 1559  
QY 62 --V-----I-D-----NN-----T-S-----68  
DB 1560 EHHVKLWKEPPIRDGSPITNYIIEVMDKDSGEFVKAIVETDSPVCKGVVKKLEGGQYKF 1619  
QY 69 ----N-----EE-----AK-----I-K-----E-----76  
DB 1620 RVRVNAKAGSPSPSEQTWNHVAKPRFLKPHIDRVNLKPVIVKTGLSISLDINIRGEPAPK 1679  
QY 77 -E-----NS-----N-K-----SQ-----83  
DB 1680 VEMFNNSSVTSDEHSVKIDNVYNTKFFVMRAQSSQSKYIIKATNEGEDEAELEVTV 1739  
QY 84 ----V-----G-----DY-----TD-----S-----89  
DB 1740 LGKPGKPGPLQVNDITKHSCKLWKEKPDGSGSPIDYIEIEKLDPHGTQWLPCKGSTP 1799  
QY 90 ----P-----V-----N-K-----NP-----KKE-----DK--V 104  
DB 1800 EAKVIGLHGKAYKFRVRAVNAKGESEDELEKPIIAKNPYDEPRPGKPEPTNNDKQDV 1859  
QY 105 ----V--Y--I-----A-----BF-----110  
DB 1860 DLANDPPKNDGAPQIKYVIQMRDKSGRAWDSATVPGDKNGTVTGVEEGHEYEFRIVA 1919  
QY 111 ----K-----D-----KE-----S-GEK-----A-IK-----121  
DB 1920 VNAKAGSPSPSVKSVIAKPRFLKPHIDRKNLQKKIMRSGQMLHIDALIKAPPAKVWT 1979  
QY 122 ---EL-SS--LK-N---T---KV---LY---133  
DB 1980 YNKTEIKTSDHIKIEENEDYKTTFFIMPVKVRADRGYIIVTAKNDSGSDTVEVELEVLCPS 2039  
QY 134 ----T-----Y-----D-- 136  
DB 2040 KPKGPLAVSNVTAETLHLKWEKPEDDGGDPIEQYLVERMDTETGRVPLTTKTPEADVT 2099  
QY 137 ----R-----I-----F-----N-----140  
DB 2100 GLTEGKEYLFRVKAIVNEEGESEPLVTDIPTAKNPFDAADTPGKQPIVDWMSGNHCCLKWR 2159  
QY 141 ----G-S-----AIET--TPD-----N-L-----151  
DB 2160 APEDDGGASITGYIVERKDPNTGKQKALETSTPDCKARVNDLIAGNKYQFRIMAVNAKAG 2219  
QY 152 ----D-----K--IK--Q-----IE--G-----159  
DB 2220 KSKPSEPDQMTAKDRFAPPKIDRNIKDITIKAGQHIRFDIKVSGEPATKVLHNAKAR 2279  
QY 160 ----IS-----S-----162  
DB 2280 LENDDSNYNIDMESYRTKLTVPISKRPHSGKYTLKAENESGRDEASFVILDKPGPBG 2339  
QY 163 -----VE-----R-----AQ-----167



Db 2340 PLRVTVHKEGCKLKMNALDGGGLPIDHYIIKMDVSGRWLPSGRFESFAELINDEP 2399  
 QY 168 ---K---V-----QPM-----MN-----HA----- 176  
 Db 2400 SHEYFRLAVANTGESEBELTGEOSVIAKNPFDEPGKGTPEAVMDKDHDVLRPPIIN 2459  
 QY 177 ---R-----K-EI---G-----V-----E----- 183  
 Db 2460 DGGSPITGVVVEKREKGTDMKIKGTETITPCIGEECKATVPTLNMENCEYFRVKAINAAG 2519  
 QY 184 -E-----A-ID-----Y-LKS-----IN---AP----- 195  
 Db 2520 PGEPSDASKPIITTKPRKLAPKIDRKNIIRTYNFKSGEPIFLDINISGEBAPDVIMQNNKS 2579  
 QY 196 ---F-----G-K---NF----- 200  
 Db 2560 VOTTSFSHIENLPYTKYINNPERKDTGLYKISAHNFYGDQDFQIINTTKPKGPEGP 2639  
 QY 201 ---D-----D-----GR-----GMV-- 206  
 Db 2640 LEVEVHNDGCKLKKKKRKDDGGEVESYLEKFPDPTGIMLPVGRSDGPEYNDGLPBG 2699  
 QY 207 --- 206  
 Db 2700 HDYKFRVAVNKEGSEPLETGSIIAKDPFVPTKPGVPEPTDWTANKVELAMPPEASD 2759  
 QY 207 --- 207  
 Db 2760 GGSPIQYIVVEKDKYSLWMEKALETNSPTPTATVOGLIEGNEYOFRVVALNKGGLSEPS 2819  
 QY 208 -S-----NI-----D---TG----- 213  
 Db 2820 DPSKFTAKPRYLAKIDRNRNITLSSGTLKLDANITGEPAPKEWKLSNYHLSGSK 2879  
 QY 214 ---T-DY-----R----- 217  
 Db 2880 NNTIETPDYTKLVRPTQSRSDGEYLVATNTSGKSVLVNVVITDKPSPNGLQISD 2939  
 QY 218 -HK----- 219  
 Db 2940 VHKEGCHLWKRRPSDDGGTPIEYFOIDKLEPETGCMIPSCRETPQVDVTGLSPGNEYKF 2999  
 QY 220 ---A-----WR-----ID--- 224  
 Db 3000 RVSAVNAEGESOPLVGDESIVAKNPFDEPGPENIKATDMDKHVDLAWTPLLIDGSGPI 3059  
 QY 225 ---D-----D----- 226  
 Db 3060 SCYIIKODKYKMERALDVPAQCKATIPDLVEGOTYKFRVSAVNAAGTGPSPSTPEPI 3119  
 QY 227 -AKA-----SM--R-----F-----KKE----- 236  
 Db 3120 IAKANNKPIIDRSLSLVEYRIKAGOSFTFDCVSGEPAPQTKMLLKKEKVSXONVKTIN 3179  
 QY 237 -D---LK-----G-----T--DK----- 243  
 Db 3180 VDYNTKLVNSATRSDSGLYTVFAENANGESADVAVKIVIDKPADPNGLKYDEINSESC 3239  
 QY 244 ---NY-----NL-----WL-----SD-----KI-----P-H--AF----- 255  
 Db 3240 TLHNNPPDDGGQPIIDNYVVEKLDGTGTGRWIPAGETDGVTLAKYGGTLPGHKYFRFRA 3299  
 QY 256 -NY-----Y-----N-----GG-KIT-- 264  
 Db 3300 KNROGTSEPLTTAQAIIAKNPFDPVTKPGTPTIKDFKDFVDLEWTRREADGSGPITGV 3359  
 QY 265 VEK-----Y-----DD-----G----- 271  
 Db 3360 VEKDKFSPDMEKCAEISDDITNAHVPLDIEGLKYEFRVAVNAKAGPSPSDATETHVAR 3419  
 QY 272 ---R---D-----Y-FD-----P-----H-G-M----- 280

Db 3420 PKNTPEKIDRNFMSDIKIKAGNVFEDVPVTEBPLPSKMDWTEGNNMINTDRVXISNEDD 3479  
 QY 281 --- 286  
 Db 3480 RTKIRILDAKRSDTGVYTLTARNINGTDRHNKVTIILDAPVPRGGLRANGVSKNSIVLR 3539  
 QY 287 ---AG-----N-----DTE----- 292  
 Db 3540 WRPPKDDGSEITHYVVEKMDNEMAMRWVPVGDCTDTEIRADNLIENHDSFRVAVNKG 3599  
 QY 293 --- 295  
 Db 3600 QSQPLTTSQPIYAKDPVSHPDKPGOPQATDWMGKHFDLEWSTPKRDDGAPISYIIKXP 3659  
 QY 296 ---K-----N-----F-----N-G-----I----- 301  
 Db 3660 KFGOMERAVALGDNCKAHVELTNGGEYFRVIAVNGGSDPSDPSSTIICKPRFLAP 3719  
 QY 302 ---D-----G---I-A--P-----NA-----Q----- 309  
 Db 3720 FFDKSLNDITVHAGKRLGWTLPLEASPRPLITWLXNGKEIGSNSRGSGGLFQNELTFEI 3779  
 QY 310 ---I---F----- 311  
 Db 3780 VSLRSDGARYTLILKNEHGSFDDASAHATVLDRESPKRPDITKITRGCHLTWNVPPD 3839  
 QY 312 ---S---Y---KM-----YSDAG-S-----G----- 322  
 Db 3840 DGGSPILHYIIKNDLSRSTWSAGMSTHIVHDVTLVRHREYLFVAVANAIGESDPLE 3899  
 QY 323 ---F---A-G-----D----- 326  
 Db 3900 AVNTIIAKNEFDEBDAPEKPIITDMDRHDIDLOMAVPSDGAIPSEYIIOKKEKSPYW 3959  
 QY 327 ---E-T-----M----- 329  
 Db 3960 TNVRHVPENKNTTITPELTGEGEYFRVIAVNOGSESESPSDIMAKPRYLPPKITTP 4019  
 QY 330 ---FHA---I-E-----D-----SI-----K- 338  
 Db 4020 LNEVRIKGLIFHTDHFIEGPAPEATWTLNSNPLSNDRSTISIGHSVHTVNCORS 4079  
 QY 339 ---H---N---VD-----VV-----SVS-S-----G--- 349  
 Db 4080 DSGIYHLLNRSSGIDGSEFELVLDLRGPRGEGMEYEITANSVTTISMKPKONGSSEI 4139  
 QY 350 ---FT-----CT-----G---L--- 355  
 Db 4140 SSVYIEKRDLTNKGGMWPAVNVYSAKYNAHVPRLLBECTWELRMAMENLQGRSDPLTSD 4199  
 QY 356 ---V---G-E-----K-----Y-----W-- 361  
 Db 4200 QPVVAKSQYTVGAPGKPELTDSDKNHITIKMKOPISNGSPIIGYDIERRDVTGRWIK 4259  
 QY 362 --- 368  
 Db 4260 INGQVPPTAEYQDORVTSNHOYQYRISAVNAAGNKTSEPSAIFMARPLREKPRYFDGL 4319  
 QY 369 ---K-AG-----IP-----AI---RALR----- 373  
 Db 4320 IGRHIVYRAGEPVNLNIPISGAPPTTIEWKRGDLKLECKRISYETNSERTLFRIDSNR 4379  
 QY 374 --- 377  
 Db 4380 RDSGYVTVANEFKDTADIEVIVVDPKSPPEGPLSTETAPDHISLHMWSPKDDGSD 4439  
 QY 378 -TG-----N-----Y-----A---TS----- 384  
 Db 4440 ITGYIIETFEFGVDWKPEVPGTCENTNFTVKNLVEGKKYVRIRAEINTYGASEALEGKPV 4499  
 QY 385 -A-S-----S---S---W---D----- 391  
 Db 4500 LAKSPFPDPCAPASQPTISAYTPNSANLEMHNPDDCGGKPIINGYIVERREGGEWIKNNY 4559

QY 392 ---L-VAN---N-----H----- 397  
Db : : : : :  
4560 PTPNTSYTNLRDYGARYFRVLAYNEAGPHGSPKSDPMTAEHQRYRDPPEPPKPDRI 4619  
QY 398 ---L---K---M---T---D---T---G--- 404  
Db : : : : :  
4620 TRNGVTLSWRPRTDGKRIKGYVEMRPKNGKDKMTVNDIPINVTYVTPSLKEGBEYS 4679  
QY 405 ---N---VT---R---T---A--- 410  
Db : : : : :  
4680 FRVVAENEGRSDPSKPSQITIEBQPNKPCWELGKVRDIVCRAGDDFSIHVPYLAFPKP 4739  
QY 411 -A-----H----- 412  
Db : : : : :  
4740 NAFWYSNDNMLDNNRVHKLHTDDAASVVVKNKSRADSGQVRLQJLKNSTGFDTAFINRV 4799  
QY 413 ---E---D---AI---A---VAS--- 420  
Db : : : : :  
4800 LDRPSPTRLRADEFSGDSLTLYWNPNDGGSALQNYIEKKEARSSTWKSYSFCTVP 4859  
QY 421 ---AKN---Q---TVE-----FDKV-N----- 432  
Db : : : : :  
4860 FVIRNLVNLKEYDFRVIAENKYGSDPANTSEPILARHPD-IDNTPGIPHIGIDSTEDS 4918  
QY 433 -I---GG-----E-----S-----FK-- 439  
Db : : : : :  
4919 ITIAWTKPHGDSGITGYIEIKRLLSDDKWTKAVHALCPDLSKIPNLNIAENAEYFRVA 4978  
QY 440 ---Y-----RN---I-G-AF----- 446  
Db : : : : :  
4979 AVNAGQAYSQSSOLIFCRPPHAPKITSDLSIRDMTVIAGDEFRIITVPYHASPRTAS 5038  
QY 447 ---FD-----KS-K-----IT-T-NE-----D- 457  
Db : : : : :  
5039 WSLNGLEVIPEGRIKFDSDYASMYNKSARDETSYITITNNKSGDTSCHVTVDVR 5098  
QY 458 ---G---T---KAP---S-----KL-----K---FV---Y- 469  
Db : : : : :  
5099 PLPPQGLNAVITPDTCTLAWKTPLDGGSPTNYVYVEKLDNGSMWKISSFVRNTHYD 5158  
QY 470 I-G---K---G-----Q---D---Q---D---LI 479  
Db : : : : :  
5159 VNGLEPHYKNFRVAENQYGLSDPLDIEPIVAKHQFTVPDEPQPKVIDWDSGNVTIL 5218  
QY 480 ---G---L-----D----- 482  
Db : : : : :  
5219 WTRPLSDGSGRIQGYIEVRDILNDSMNAYDIIKTKYQLYNLINGSEYEFRIKAKNA 5278  
QY 483 ---LR---GK-I---A-----V-----M-D---RI----- 493  
Db : : : : :  
5279 AGLSKPSSPSLRFKLKGRFTVPSPPGAPQVTRVGKYYVDLKWKEPLRDGSGRITGYIIER 5338  
QY 494 ---YT---K-DLK---N----- 500  
Db : : : : :  
5339 RDIGAVVWKNDYVNLQTEYTMNLIEMGDYFRFVAVNSAGRSEPSLCTWPIKCEVL 5398  
QY 501 ---A-F-K-----K---A-----M--D-K--- 508  
Db : : : : :  
5399 GGGKPDWITRLQDKVAPFGKDYTLQCAASGKPSPTARLNGKEIQMNGRMTCDSDKGV 5458  
QY 509 ---G---A-----R----- 511  
Db : : : : :  
5459 FRLHISNVQTDGDDYTCEAMNSLGFVNTSGYLKIGSPPIINRCPSELKLPEDNSKIKI 5518  
QY 512 ---AIW---VV---N---T---VN---Y---YNR---D--- 525  
Db : : : : :  
5519 FYSGQOPLTVILUKKNEVICDSNDDTHVKVNIFFDYVAIYIRNIVKSDGGPYQIEFTNES 5578  
QY 526 ---NW----- 527  
Db : : : : :  
5579 GSATGEFYVHITGMPAPTGPNGISYINKNSCMLNWRPPSYDGLKLVSHYVIERKDVSSP 5638

QY 528 ---T-----E-L-P-AM----- 533  
Db : : : : :  
5639 HWITVSSTCKDTAFNVOGLIENQEIYIFRMAVNEGMGPPLLEGLNPIRAKPIDPPSPPG 5698  
QY 534 ---G---YE-----A-----D----- 538  
Db : : : : :  
5699 SPQITEIGGDFVHLEWKEKPEDGGAHQYIQQWIDKREVSNTQWRVATNATCAANQINCINL 5758  
QY 539 -EG---TKS---Q---V-----P--- 546  
Db : : : : :  
5759 IEGROYEFRIPAQNVAGLSTESSASQAVKIIDPOAASPLIVKPLRDANCIONHNAQFTC 5818  
QY 547 ---S---IS---GD---D-G---V----- 554  
Db : : : : :  
5819 TINGVPKPTISWYKGAREISNGARYHMYSEGDHFLNINOVFGEDADBYVCRAVNAKAGAK 5878  
QY 555 ---KL-----N-----W-----N----- 558  
Db : : : : :  
5879 STRATLAIMTAPKLVPPRFRDTAYFDKGENVVIKIPFTGPKPRIHWRDGENIESGGH 5938  
QY 559 ---MI-----N-----PDK-----K--- 565  
Db : : : : :  
5939 YTVEVKERHAVLIIRDGSHLDGSPYRITAENELGSDTAIQVQISDRPDPFRPLIESIG 5998  
QY 566 TE-----VKR-----NN-----KE-DF-- 576  
Db : : : : :  
5999 TESLSWKAPWDGCSGITNYVVERREHPLUSSWIRVGNTRFTSMVSGLTPGKEYDPRI 6059  
QY 577 ---K---D---KL----- 580  
Db : : : : :  
6059 FADNVYGRSDASDTSLIKTKESVKKKPIERKWEIDANGRLRGKADGPKVDYDYSYVFDI 6118  
QY 581 ---E-Q-Y-Y-----P----- 585  
Db : : : : :  
6119 YSKEVPOPEISQGSVYDRYDILBEIGTGAFGVHVHRCRSTGNIFAAKFIPVSHSVEKD 6178  
QY 586 ---ID-----M---E-S-F-----N 592  
Db : : : : :  
6179 LIRREIDIMNQLHHQKLNLDHAFEDDEDEMIILIEFLSGGELFERITAEGVYMTAEVIN 6238  
QY 593 ---SN---KP-----NV---G---D-E-K--- 602  
Db : : : : :  
6239 YMRQICEGIRHMQNIHLDIKPENIMCQTRSTNVKLIDFGLATRLDPNEVVKITGT 6298  
QY 603 ---EI-----DF----- 606  
Db : : : : :  
6299 AEFAAPEIVNREPVGFTYDMMATGVLVSYLLSGLSPFAGDNDVOTLKNVKACDWDVDVES 6358  
QY 607 ---K-F---A-----P---D-T---D----- 613  
Db : : : : :  
6359 PKYISEEAKDFIRKLLVRNKEKRMTHAECLLHPWLTDGHSAMKQEIINRDRYLAYREKURR 6418  
QY 614 K---E---L-Y---K-E-E---D---I----- 621  
Db : : : : :  
6419 KYEDFERLLPIGRLESEYSSLRKLLMEKYKTHDAVDRRQAPRFRVIRPSQFCVQSGOV 6478  
QY 622 ---I-V-P-----A--- 625  
Db : : : : :  
6479 KFYCRCAIATPTLTWSHNNIELRQSVKFMKRYVGGDYFYIINRVKLLDDRGEYIIRAENH 6538  
QY 626 -GS-----T-S-----W----- 630  
Db : : : : :  
6539 YGSREEVFLNVQPLPKBPQRYRTESTPVRREPLPYTTFWEESETPASFTFLLRPRVMQ 6598  
QY 631 ---G---P-----R-----ID----- 635  
Db : : : : :  
6599 ARDTCKLLCCLSGKPVNVRWYKDGRELSKVEYAMTHSDGVVTWEIIDCKPSDSKYSCK 6658  
QY 636 ---L---LL---KP----- 640  
Db : : : : :  
6659 ATNCHGTDETDCVIVIEGEMWTPQOQLAHNFLYSGDKRYEQRIKPAPLPIVTSROYTS 6718  
QY 641 ---D---VS-----APG----- 646

```

Db      6719  SSVNTSEPGCDKVVNSNSGSGISNKKVANSLSLOAGSPSRSRATKEILPPDDSLM 6778
Qy      647  -----K-----I-----K----- 650
Db      6779  CKPEFTKPLDLTTHDEQLILTCYKGDPEPQISWKSGLSSSDILRLRYNGIATL 6838
Qy      651  -----S-----T-----L-----N-----V-----I----- 656
Db      6839  TINEFPEDEGVICTATNSVGAETKCKLTIQPLDKINKRKYNAGNARKIVSHLES 6898
Qy      657  -----N-----G-----K-----S-----T-----Y-----G----- 663
Db      6899  FVRDGAVALNLCRIIGAQHPDVVWLHNKKEIKPSKDFQYNEANIYRLQIAEIFPEDG 6958
Qy      664  Y 664
Db      6959  Y 6959

```

## RESULT 9

ABG51536 ID ABG51536 standard; Peptide; 5701 AA.

XX AC ABG51536;  
 XX DT 25-FEB-2003 (first entry)  
 XX DE Human liver peptide, SEQ ID No 30184.

XX KM Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;  
 XX KM hypercholesterolaemia; coronary heart disease.  
 XX OS Homo sapiens.

XX PN MO200157273-A2.  
 XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00664.  
 XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.  
 XX PR 30-JUN-2000; 2000US-0608408.  
 XX PR 03-AUG-2000; 2000US-0632366.  
 XX PR 21-SEP-2000; 2000US-0234687.  
 XX PR 27-SEP-2000; 2000US-0236359.  
 XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-488898/53.

XX PT Human genome-derived single exon nucleic acid probes useful for  
 XX PT analysing gene expression in human adult liver -

XX PS Claim 27; SEQ ID No 30184; 658bp; English.

XX CC The invention relates to a single exon nucleic acid probe (SENP) (I) for  
 XX CC measuring human gene expression in a sample derived from human adult  
 XX CC liver, comprising one of 11109 defined nucleotide sequences given in the  
 XX CC specification (or complements/ fragments). The probe hybridises at high  
 XX CC stringency to a nucleic acid molecule expressed in the human adult  
 XX CC liver. (II) may be used for predicting, measuring and displaying gene  
 XX CC expression in samples derived from human adult liver. The genes  
 XX CC identified may be involved in genetic liver diseases such as cirrhosis,  
 XX CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which  
 XX CC is associated with coronary heart disease. ABG47348-ABG5930 represent  
 XX CC human liver single exon encoded peptides of the invention.  
 XX CC Note: The sequence information for this patent does not appear in the  
 XX CC printed specification but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX SO Sequence 5701 AA;

Query Match 69.8%; Score 3070.2; DB 22; Length 5701;  
 Best Local Similarity 10.3%; Pred. No. 7e-44;  
 Matches 578; Conservative 68; Mismatches 14; Indels 4958; Gaps 495;

```

Qy      1  Y-----P-----V-----LA-----D-----T-----S----- 9
Db      73  YNFRISAINDAGCEPAVIPPVEIVEREMAPDELDALRRLTVVRAGLSIRIFVPIKGR 132
Qy      10  -----S-----S-----E-----D-----A----- 14
Db      133  PAPEVTXKNINLNKKNANIENTESFTLLIIPECNRYDTCKFVWTIENPAGKXSGFVNR 192
Qy      15  -----LN-----I-----S-----D-----K-----E----- 21
Db      193  VLDTPGFVLNLRPTDITKDSVTLHMDPLIDGSRITNYIVEKREATRKSYSTATTKCHK 252
Qy      22  -----KV-----AEN-----K-----E----- 28
Db      253  CTYKVTGLSBCGEYFFRVMAENEGYIGEPTEETTEPPVASEAPSPDLSNIMDTKSTVSL 312
Qy      29  -----KH-----E-----N-----I-----H----- 34
Db      313  AWPKKXIDGSGSKITGYIEAQRKSGDQWHTITTYKGLCEVVRNLTEGEEYTFQVMVNSA 372
Qy      35  --SA-----M----- 37
Db      373  GRSAPRESRPVIVEQMLPELDLRCIYQKLVIAKAGDNIKVEIPVLRPKPTVTKKGD 432
Qy      38  -----ET-----S-----O----- 41
Db      433  QILKOTGVNFEFTATSTILININECVRSDSGPPLTARNIVEGVDTITQVHDI PCGPPT 492
Qy      42  -----DF----- 44
Db      493  GPIKFEVSDVFYFMSMDPENOGVPISNVVMEMRQTDSTTWELATVIRTYYKATRL 552
Qy      45  --E-----K-----TAV----- 50
Db      553  TTGLEQYFRVAQNRVGVGPGITSACIVANVPKVPGPPTQVATVATKOSMTSMHEPL 612
Qy      51  -----I-----KE-----K-----E-----V----- 57
Db      613  SDGSPILGYHVERKENGILMOTVSKALVPGNIFKSSGLTDGIAYEFVIAENMAGSK 672
Qy      58  -SK-----N-----P-----V-----I-----D----- 64
Db      673  PSKPSPEMLALDPIDPGKPPVPLNITRHTVTLKMAKEFYGGFKITSTIVEKRLPNGRW 732
Qy      65  -----N-----T-----S-----E-----EA-KIK- 75
Db      733  LKANFSILNEBEPFVSGLTEDAAYEFVIAKNAAGAISPEPEBDATCRDQVAPKIKV 792
Qy      76  -----E-----E-----E-----N-SNK- 81
Db      793  DVKFKDVTILKAGAPRLADVSGRPPTMEWSKQKLEBQAKLEIKIADPSTNLVVKD 852
Qy      82  S-Q-GDYT-----D-----S-----F----- 90
Db      853  STRDSCAVYLLTATNPGCFAKHITNVKVLDRPGPEGPLAVTEVTSEKCVLSWPPPLDDG 912
Qy      91  -----V-----N-----K-----N-----T----- 95
Db      913  GAKIDHYIVQRETSRLAMTNVASEVQVTKLVTKLKNGEYIPRVMAVNVYGVGPLES 972
Qy      96  E-----N-----PK-----K-----E-----DK-----V----- 104
Db      973  EPVLAVPYGPDPKPNPEVTTITKDSMVVVCWCHPDGSGSEIINYIVERDKAGGWRK 1032
Qy      105  -----V-----Y-----I-----AE-----FK----- 111

```

Db	1033	CNKKTLTLRLRYKVSGLTEGHEYEFPRMAENAGISAPSPSPFYKACDVTFKPGPPGNPR	1092
Qy	112	-----D-----K-----	113
Db	1093	VLDTSRSSISIAWNKPIYDGGSEITGYMVEIALPEDEWQIVTPPAGLKATSYTITGLTE	1152
Qy	114	E-----S--GE-----KA-----	119
Db	1153	NOEKIRIYAMNSEGLGSPALVPGTPKAEADRLPPEIBLDADLRKVITIRACCTLRFPV	1212
Qy	120	IK-----E--L-----SS-----L-----K-----	127
Db	1213	IKGRPAPEVKWARDHGESLDKASTESTSYTLLIVGNVNRFDGKXIILTVENSSGSKSAF	1272
Qy	128	N-----TK--V-----L-----Y-----T-----Y-----	135
Db	1273	VNVRVLDTPGPPQDLKVKKEVTKTSVTLTWDPPLDGGSKIKNYIVKEKRESTRKAYSTVAT	1332
Qy	136	-----DRI-----F-----N-----G-----S--A-----	143
Db	1333	NCHKTSMVKDLOEGCSYFRLAENEGYIGLPAETAESVKASERPLPPGKITLMDVTRN	1392
Qy	144	-----I-----I-----E-T-T-----	147
Db	1393	SVLSWEKPEHDGGSRIILGYIVEMOTKGDGSKWATCATVKVTEATITGLIQEEYSFRVSA	1452
Qy	148	-----P-----D-----N-----L-----D-K-----I-----	154
Db	1453	QNEKISDPRQLSPVIAKDLVIPAPKLLFNTFTVLAGEDLKVDVDPFGRPTPAVTHWK	1512
Qy	155	-----KQ-----	158
Db	1513	DNVPLKOTTRVNAESTENSLTTIKDAREDVGHVYVVKLTNSAGAEIETLNVILDKPGP	1572
Qy	159	-----G-----I-----SS-----VE-----R-----A-----	166
Db	1573	PTGPVKMEVTAADSIITLSWGPBPKYDGGSSINNYIVEKRDSTTTTQWIVSATVARTIIKAC	1632
Qy	167	-----Q-----K-----V-Q-----P--M-----M-----N-----	174
Db	1633	RLKTGCEQFRIAENRYGKSTYLNSEPTVAQYPPKVPKPGPTPVVTLSSRDSMEVQWNE	1692
Qy	175	-----H-ARKE-----I-----CWE-----	183
Db	1693	PISDGSRVIGYHLERKERNISILWYKLNKTPIPOTKFKTTGLEGEVEYEFVRVSAENIVGI	1752
Qy	184	-----EAI-----	189
Db	1753	GPKSVSECYVARDPCDPPGRPEALIVTRNSVTLOWKKPTVDGGSKITGYIVEKKELPEG	1812
Qy	190	-----K-S-----I--NA-----	194
Db	1813	RWMKASFTNIIDTHFEVTGLVEDHRYEFRVRIARNAAGVFSPESESTGAIARDEVDPRI	1872
Qy	195	-----P-----F-----GK-----N-----F-----	200
Db	1873	SMDPKYKDTIVVHAGESFKVDADYIGKPIPTQWIKGQELSNTARLEIKSTDFATSLSV	1932
Qy	201	-----D-----G-----R-----G-MVIS-----	208
Db	1933	KDAVRVDSGNVILKAKNAVAGERSVTNVVKVLDRLPGPPGPPVIVSGVTAETKCTLANWKPLOQ	1992
Qy	209	-----N-I-----D-----T-----G-----	213
Db	1993	DGSDIINYIIVERRETSRLVWTVVDANVQTLSCVKTKLEGNEXYTFRIMAVNKYVGEPPL	2052
Qy	214	-----T-----D-----Y-----	216
Db	2053	ESEPVAKNPFVDPAPKAEVTTVTKDSMIVWWRPASDGGSEITLGVLEKRDKEGIRW	2112
Qy	217	-----R-----H-----KA-----	220
Db	2113	TRCHKRLIGELRLRVTLGIENHDYEFVRVSAENAGLSBSPSPSYQKACDPYKPGPPNN	2172
Qy	221	-----M-----	221
Db	2173	PKVIDITRSSVFLSWSKPIYDGGCBIOGYIVEKCDVSVGEWMTCTPPTGINKNTINEVEKL	2232
Qy	222	-----RI-----D-D-D--AK-----A--SMR--F-----	233
Db	2233	LEKHEYNFRIICAINKAGVGEHADVPGPPIVEBEKLEAPDIDLDLELRKIINIRAGGSLRUF	2292
Qy	234	-----K--K--E-----D-----L-----K-----GT--	241
Db	2293	VPIKGRPTPEVKWGVKVDGEIRDAALIDVTSSFTSLVDNVNRYDSGKYTLTLENSSGTKS	2352
Qy	242	-----W-----L-----S-----D-----KNY-----	245
Db	2353	AFVTVRVLDTPSPVNLKVTEITKDSVSITWEPPLDGGSKIKNYIVEKREATRKSAAV	2412
Qy	246	-----W-----L-----S-----D-----KI-----	251
Db	2413	VTNCHKNWKIDQLOEGCSYFVRVTAENEYIGLPAQTADPIKVAEVPQPGKITVDDVT	2472
Qy	252	-----P-H-----A-----F--	255
Db	2473	RNSVLSWTKPEHDGSGKIIQYIVEMQAKHSEKSECARVKSLOAVITNLTOGEEVLFV	2532
Qy	256	-----N-----Y-----	258
Db	2533	VAVNEKGRSDPRSLAVPIVAKDLVIEPDVKPAFSSYSVQVGQDLKIEVPISGRPKPTIW	2592
Qy	259	-----N-----GCK--ITV-----E-----	266
Db	2593	TKDGLPLKOTTRINVTDSLDLTLSIKETHKDDGGQYGITVANVVGQKTASIEITLDKP	2652
Qy	267	-----KYDD-----G-----RD-----	273
Db	2653	DPPKGVKFDVDSABESITLSWNPPLYTGCGQITNIVOKRDTTTTVDVVSATVARTLLK	2712
Qy	274	-----Y--F--DP-----	277
Db	2713	VTKLKTGYOFRIPAENRYGQSFALSDPIVAQYPPKEGPPGTPPFATAISKDSMVIQW	2772
Qy	278	-----GMH-----I-----A-----GI-----LA-----	287
Db	2773	HEPVNNGSPVIGYHLERKERNISILWTKNKTIIHDTQFKQNLBEGIEYEFVRVYAENTV	2832
Qy	288	-----G--N-----D--T-----E-Q--D--I-----K-----	296
Db	2833	GVGKASKNSECYVARDPCDPPCTPEPIWKNRNEITLQWTKPVYDGSMTGYIVEKRDLP	2892
Qy	297	-----N-----F-----N-G-I--D--G-I-A-----	305
Db	2893	DGRWMKASFTNVIEQFTVSGLTEDQRYEFRVIAKNAAGAIKSPSDSTGPIAKDEVELP	2952
Qy	306	-----P-----NA-----	308
Db	2953	RISMDPKFRDITVNVAGETFRLEADVHGKPLPTIEWLRGDKIEBSARCEIKNTDFKALL	3012
Qy	309	-----Q-I-----F-----	311
Db	3013	IVKDAIRIDGGQYILRASNVAGSKFPVNVKVLDRPGPEGPVQVGTVSEKSLTWSPP	3072
Qy	312	-----S-Y-----K-----M-----Y-----	316
Db	3073	LQDGGDISHYVVEKRETSRLAWTVVASEVNTLSKVLKLEGEVYFRIMAVNKYVGVE	3132
Qy	317	-----SDAG-----	322
Db	3133	PLESAPVLMKNPFLVLPGPPKSLVETNIAKDSMTVCNWRPDSGGSEIIGYIVEKDRSGI	3192
Qy	323	-----F-----AG-----	330
Db	3193	RWIKCNKRITDLRLRVTLGTLDEHDEYEFVRVSAENAGVGEPSPATVYVYKACDPV-FKPGP	3251

QY 331 ---H-----A-----I-----E-----D-----S-----I 337  
 Db 3252 PNNNAIIVDTKNSITLWAKMCPYIDGSELGVVLEICADEEEMQIVTPQICLARTTREI 3311  
 QY 338 -K--H-----N-----V-----D-----V-----S 345  
 Db 3312 SKLTHEOEYKIRVCAIANKVGLGEATSVGPVAKPEDKLEAPBLDLSLRKGIIVRAGSA 3371  
 QY 346 ---S-----G-----F-----G-----T-----G-----L----- 355  
 Db 3372 RIHIFKGRPIPEITWISREGEFTDKVOIEKGVNTQJLSIDNCDRNDAGKYLKLENSSG 3431  
 QY 356 ---V-----G-----E-----K-----Y-----W-----A-----I-----R----- 365  
 Db 3442 SKSAFVTVKVLDPGPBPNLAVKEVRKDSAPFLWMEPRIIDGSAKKNVVIDKRESTRKAY 3491  
 QY 366 A-----L-----R-----KA-----GIP-----MV----- 375  
 Db 3492 ANVSSKSKTSFKVENTLEGAIYFRVMAENEFVGVPETVDVAKAAEPSPGKVTLT 3551  
 QY 376 -V-----AT-----G-----N-----Y-----381  
 Db 3552 DVSQTSASLWMEKPEHDGSRVLGVVENQPKTEKMSIYAESKVCNAVVTGLSSGOEYO 3611  
 QY 382 ---A-----S-----A-----S-----S-----S----- 388  
 Db 3612 PRVYAKNEKGSDDPRLVGPVIAKDLTLOPSLKLFPNTYSIOAGEDLKEIFVIGRPRPN 3671  
 QY 389 -SW--D--L-----VAN-----N-----HLK-----MT----- 401  
 Db 3672 ISWVADGEPLKQOTTRV-INVEETATSTVLHIKEGNKDDFGKYVTATNSAGTATENLSVIV 3730  
 QY 402 ---D-----TG-----N-----VT-----R-----TAA--H----- 412  
 Db 3731 LEKPGPVGPRVDFEVSADPFIWISMEPRAYTGCOISNYIWEKRDITTTTWMVSATVAR 3790  
 QY 413 ---E-----D-----A-----IA-----VAS--AK----- 422  
 Db 3791 TTIKITLKTGTETGYQFRIAEKRRYKSAFLDSKAVIYOIFKEKPGPGPTVTSISKDOM 3850  
 QY 423 ---N-----O-----T-----V-----EF-----D-----KV----- 431  
 Db 3851 LVQWHEPVDGKTIIIGYHLEQEKNSILWVKLNTPIODTKFTTGLDEGLEIEFKSA 3910  
 QY 432 -NI--G--GE--S--F-----K-----Y-----440  
 Db 3911 ENIVIGIKPSKVSCEGFVARDPCDPGRPEAIVITRNNTLKKWKAPAYDGSKITGIYEVK 3970  
 QY 441 ---R-----NI-----G-----A-----F-----446  
 Db 3971 KDLPGRMWKAFTNVLTEFTVSGLVEDQRYEFRIYANNAAGNFSPPSDSGAITARDE 4030  
 QY 447 ---F-----D-----K-----SK-----I 452  
 Db 4031 IDAPNASLDPKXVDIVVHAGETFVLEADIRCKPIPDVWVMSKDKLELETAAMEIKSTI 4090  
 QY 453 ---TT-----N-----E-----D-----G-----T-----K----- 460  
 Db 4091 QKTTLVVKDCIRTDGQYILKLSNVGKTSIPITVAVLDRPGPREGPLKVTVTAEKCYL 4150  
 QY 461 A-----P-----S-----K-----L-----K-----F-----V----- 468  
 Db 4151 ANNPLQDGGANISHYIIEKRETSRLSWTQSTEVQALNYKVTKLLPGNEVIFRVAWVK 4210  
 QY 469 Y--IG-----K-----G-----Q-----D-----Q-----D-----L----- 479  
 Db 4211 YGIGBLBSGVPYACNPKPBPSPSTPEVSATITKOSMVTWAPVDDGTEIEGYILEKR 4270  
 QY 480 ---G-----L-----DLR-----G-----KIA-----488  
 Db 4271 DKEGVMTKCNKKTLLDLRLRVLTGLTEGHSYEFRVAAENAGVGEPSSESVFYACDALY 4330

QY 489 ---VMD--R-----IY-----TK-----D-----497  
 Db 4331 PPGPSPNPKVTDTRSRSVSLAWSKPIYDGAHPVKGVVEVKEAADADEWTTCTPTGLGK 4390  
 QY 498 ---LK--NA--F-----KKA-----MD-----507  
 Db 4391 QFTVTKLENTENYFRICAINSEGVEBPATLPGSVVAOERIEPPEILDADLRKVVURA 4450  
 QY 508 ---KG-----A-----RA--I-----M-----514  
 Db 4451 SATRLFPVTIKGRPEPEVKEKAGILTRQOIEVTSSFTWLVINDVTRFDSGRYNTLE 4510  
 QY 515 ---V-----VN-----517  
 Db 4511 NNSGSKTAFVAVVRVLDSAPAVNLTIREVKDVS TLSMEPPLIDGAKITNYIVEKRETT 4570  
 QY 518 ---TV-----N-----YY-----N-----R 524  
 Db 4571 RKAVATTINNCTKTFRIENLOEGCSYFRVLASNEVGIGLPAETTEPVKVEBPPLPGR 4630  
 QY 525 ---D--N-----W-----T-----528  
 Db 4631 VTLVDVTRNTATIKWEKPESDGSKITGYVEMOTKSEKMSCTOVKTLTATISGLTAG 4690  
 QY 529 ---ELP-----A-----MG--534  
 Db 4691 EBYFRVAANVEKGRSDPRQLGVPVIAARDIEIKPSVELPFTFVNVKAREQLKIDVPFKGR 4750  
 QY 535 ---YE-----A-----537  
 Db 4751 PQATVNMKDGOTLKETTRVAVSSKTVTSLSIKEASKEDVGYELCVSNAGSITVBIT 4810  
 QY 538 ---DE-----G-----540  
 Db 4811 IIVDRPGPPGPIRIDEVSCDSITISNMPPEYDGCQISNYIWEKETTSTTHIVSOAV 4870  
 QY 541 ---T--K-----S--Q-----V-----FS-----547  
 Db 4871 ARTSIKIVRLTGTSEYQFRVCAENRYKSSYSSESAVAEYFPSPGPGTPKVVHATKS 4930  
 QY 548 ---I-----SG-----550  
 Db 4931 TMLVTWQPVNDGSGRVIYHLEVERSSILMSKANKILADTQWKGVLDEGLMYEYRV 4990  
 QY 551 ---D-----DG-----V 554  
 Db 4991 YAEINAGIGKSCSKCEPVAPADPCDPGPQPEVNTITRKSVALKWSKPHYDGAKITGYIV 5050  
 QY 555 ---KL--W-----N-----558  
 Db 5051 ERRELPGDRWLKCNVTNIOETTFEYTELTDORIEFRVAPANAADSVSEPSSESGPIIVK 5110  
 QY 559 ---M-----IN-----P-----D-----KK--TE-----567  
 Db 5111 DQVEPRVMDVKFRDVIIVKAGEVYLKINADIAGRPPLVISWADGIEIEBRAREIIST 5170  
 QY 568 ---YK-----R-----N-----N-----K--573  
 Db 5171 DNHTLLTVKDCIRRDGQYVLTLLKNVACTRSVAVNCKVLDKPRGPAGLEINGLTAEKCS 5230  
 QY 574 ---E-----DFKD-----K-----L-----E--O-----Y-----583  
 Db 5231 LSWGRPQDGGADI-DYIIVEKREISHLAWITCEGELQMTCKYTKLLKGNBYIRVYGV 5289  
 QY 584 ---Y--P-----I--D-----M-----ES-----F-----591  
 Db 5290 NKYGVBESLBSVAIKALDPFTVPSPPTSLBITSVTKESMTLCWSPRSDGSGEISGYIIE 5349  
 QY 592 ---NS-----N--KP--N--V-----G--D--E-----K-----E-----I--D--605  
 Db 5350 RREKNSLRWVAVNKKPVVDLRVKSTGLREGEYEYRVYAENAGLSLPSSETSPLIRADBP 5409  
 QY 606 -F-----K--F--AP-----DTD--K-----614

Db 5410 VFLPSPSPKPIVDSKTTITIAWPLFDGGAPITGYTVVEYKSDDTWKTSIQSLRGT 5469  
Qy 615 E-----L-----Y-----|||  
Db 5470 EYISGLTTGAEYVPRVKNVKGASDPDSDDPQIAKEREEREPPLFDIDSEMRKTLIVKA 5529  
Qy 626 G-S-T-S-----WG-P-----R-----I-----D----- 635  
Db 5530 GASFTMTVPFRGRPVNVLWSKPDPTDLTRAYVDTTDSRTSLTIENANRNDGKYTLTIQ 5589  
Qy 636 -L-----L-----K-----DV-----SA-----P-----G-----KN-----I----- 649  
Db 5590 NVLSAASLTLLVVKVLDTPGPPTNITVQDVTKESAVLSWDVPENDGGAPVKNYHIEKREAS 5649  
Qy 650 -K-----S-T-----LN-----VIN-----GKSTY-----G-----Y 664  
Db 5650 KKAWSVTNNCNRLSYKVTNLQEG-AIYYFRVSGGENEF 5686

RESULT 10  
ABB36684  
ID ABB36684 standard; Peptide; 5701 AA.  
XX ABB36684;  
AC ABB36684;  
XX  
DT 04-FEB-2002 (first entry)  
DE Peptide #4190 encoded by human foetal liver single exon probe.  
XX Human; foetal liver; gene expression; single exon nucleic acid probe.  
XX Homo sapiens.  
OS  
XX  
PN WO200157277-A2.  
XX  
PD 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US00669.  
XX  
XX 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;  
PI WPI; 2001-483447/52.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human fetal liver -  
XX  
XX Claim 27; SEQ ID NO 29319; 639pp + sequence listing; English.  
XX  
XX The invention relates to a single exon nucleic acid probe for  
CC measuring human gene expression in a sample derived from human foetal  
CC liver. The single exon nucleic acid probes may be used for predicting,  
CC measuring and displaying gene expression in samples derived from human  
CC fetal liver. The present sequence is a peptide encoded by a single exon  
CC nucleic acid probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 5701 AA;

Query Match 69.8%; Score 3070.2; DB 22; Length 5701;  
Best Local Similarity 10.3%; Pred. No. 7e-44;

Matches 578; Conservative 68; Mismatches 14; Indels 4958; Gaps 495;  
Qy 1 Y-----P-----V-----LA-D-----T-----S----- 9  
Db 73 YNFRISAINDAGVEPAVDPDVEIVEREMAPDFELDAELRRLTVVRAGLSIRIFVPIKGR 132  
Qy 10 -----S-S-----E-----D-----A----- 14  
Db 133 PAPEVTWKDINILKNRANIENTESFTLLIPECNRYDTGKFMVTIENPACKKSGFVNVR 192  
Qy 15 -----LN-----I-----S-----D-----K-E----- 21  
Db 193 VLDTPGVNLRLPTDITKDSVTLLHWDLPIDGGSRIITNIVYKREATKRSYSTATTCKCHK 252  
Qy 22 -----KV-----AEN-----K-E----- 28  
Db 253 CTYKVTGLSEGCYFFRVMANEYIGETETTEPVPKASEAPSPDLSINIMDITKSTVSL 312  
Qy 29 -----KH-----E-N-----I-----H----- 34  
Db 313 AWPKPHDGGSKITGYVIEAQRKGSQDWTHTTVKGLGVNRNLTGEYTFQMAVNNSA 372  
Qy 35 --SA-----M----- 37  
Db 373 GRSAPRESRPVIVKEQTMPLDPLDGIYQKLVIAKAGDNKIVPLGRPKPTVTKKGD 432  
Qy 38 -----ET-----S-----Q----- 41  
Db 433 QILKQTVNFTTATSTILNINECVRSDSGPYPLTARNIVGEVGDVITIQVHDIPGPT 492  
Qy 42 -----DF-----K----- 44  
Db 493 GPIKFEVSSDFVTFSWDPPENDGGVPSNVVEMRQDSTTWVELATTVIRTYKATRL 552  
Qy 45 -----E-----K-----K-----TAV----- 50  
Db 553 TTGLEVQFRVKAQNRVGVPGITSACIVANYPVKVPGPPGTPQVTAVKDSMTISWHEPL 612  
Qy 51 -----I-----KE-----K-----E-VV----- 57  
Db 613 SDGSGPILGYHVERKERNILMOTVSKALVPGNIFKSSGLTDGIAEYFRVIAENMAGSK 672  
Qy 58 -SK-----N-----P-----V-----I-----D----- 64  
Db 673 PSKPSEPMALDPIIDPGKVPPLNITRHTVTLKWAKEPYTGGFKITSYIVKRDLPNGRW 732  
Qy 65 -----N-----T-S-----N-----E-----EA-KIK- 75  
Db 733 LKANFSNILENFTVSGLTEDAAYEPRVIAKNAAGAISSPSEPSDAITCRDDVEAPKIV 792  
Qy 76 -----E-----E-----N-SNK- 81  
Db 793 DVKFKDTVILKAGEAPRLEADVSGRPPTMEWSKDGKELEGTALEIKIADFSTNLVNDK 852  
Qy 82 S--Q--GDYT-----D-----S-F----- 90  
Db 853 STRDSCAYTLTATNPGGFAKHIFNVKVLORPGPEGLAVTEVTSEKCVLSWPPPLDDG 912  
Qy 91 -----V-----N-----K-----T 95  
Db 913 GAKIDHIVQKRETSRLAWTNVASEVQTKLVTKLLKNGEYIFRVMAVNKYGVGEPLS 972  
Qy 96 E-----N-----PK-----K-----E-DK-----V- 104  
Db 973 EPVLAVNYPGPPDPKPNPEVTTITKDSMVVCMWGHPSDGGSEIINYIVERDKAGQRWIK 1032  
Qy 105 -----V-----Y-----I-AE-----FK----- 111  
Db 1033 CNKKTLDLRYKVSGLTGEGHEYBPRMAENAGISAPSPFPFYKACDTVPKPGPNPR 1092  
Qy 112 -----D-----K----- 113  
Db 1093 VLDTSRSSISIANWKPIYDGGSEITGYMVEIALPEDEWQIVTPPAGUKATSYTITGLTE 1152

[illegible][illegible]

Db 3312 SKLTEHQEYKIRVCALNKVGLGEATSVPGTVRPEDKLEAPELDLDSLELRKGIWVRAGGSA 3371  
QY 346 ---VS-S---G-FT---G---T---G---L---355  
Db 3372 RIHIPFKGRPTTEITWSREEGETDKVOIEKGWNYTQLSIDNCDNRNDAGKIILKLENSSG 3431  
QY 356 ---V---G---E-K---Y-WQ---A---I---R---365  
Db 3432 SKSAFVTVKVLDPGPPQNLAVKVRKDSAFVWEPPIIDGGAKVKNYVIDKRESTRKAY 3491  
QY 366 A---L---R-KA---GIP---MV---375  
Db 3492 ANVSSKCSKTSKVENLTGEGAIYFRVMAENEFVGVPVETVDVAKAABPPPPGKVTLT 3551  
QY 376 -V---AT---G---N---Y-381  
Db 3552 DVSQTSASLWMEKPEHDGGSVGLGVVEMQPKTEKWSIVAESKVCNAVVTGLSSGQEQY 3611  
QY 382 ---A-T---S---S---S---388  
Db 3612 FRVKAYNEKGS DPRVLGVPVIAKDLTIQPSLKLPTNTYSIOAGEDLKIEIPVIGRPREN 3671  
QY 389 -SW--D--L---VAN---N--HLK---MT---401  
Db 3672 ISWVKDGEPLKQTRV-NVEETATSTVLHIKEGNKDDPGKYVTATNSAGTATENLSVIV 3730  
QY 402 ---D---D---TG---N-VT-R---TAA-H---412  
Db 3731 LEKPGPPVGPVRFDEVSADFVISMPEPAYTGCOISNVIKRDTTTTTHMVSATVAR 3790  
QY 413 ---E---D-A-IA---VAS-AK---422  
Db 3791 TTIKITKLTGETYOFRIFAENRYGKSAPLDSKAVIQYFPKEPGPPGTFVTSIKDQM 3850  
QY 423 ---N---O---T-V--EF---KV--431  
Db 3851 LVQWHEPVNDGTTKIIGYHLEQEKNSILWVINKLTPIODTKFTTGLDEGLEYEYFKVSA 3910  
QY 432 -NI-G-GE---S--F---K--Y---440  
Db 3911 ENIVGIGKPSKVSECFVARDPCDPPGRPEAIVITRNVTWKWKPAYDGSKITGIVBEK 3970  
QY 441 ---R---NI---G---A--F---446  
Db 3971 KDLPGRMWKASFTNVLETEFTVSGLVEDQRYEYFRVIARNAAGNFSEPSDSSGAITARDE 4030  
QY 447 ---F---D--K---SK---I452  
Db 4031 IDAPNASLDPKYKDVIVVHAGETFVLEADIRGKPIPDVVWSKDGKELEBETAARMEIKSTY 4090  
QY 453 --TT-----N-----E--D-----G-T-K---460  
Db 4091 OKTTLVVKDCIRTDGQYILKLSNVGGTKSIPITVKVLDRLPGPPEGLKVTGTAEKCYL 4150  
QY 461 A--P-----S---K-----L---K-----F---V---468  
Db 4151 ANWPPLQDGGANISHYIEKRETSRLSWTQVSTEQALNYKVTKLLPGNEYIFRVMAVVK 4210  
QY 469 Y-IG-----K-G--Q-----D-----Q--D-----LI---479  
Db 4211 YGIGLEBESGPVTACNPKYKPPPPSTPEVSAITKDSMVVTWARPVDVDDGGTEIEGYILEKR 4270  
QY 480 ---G---L-DLR-----G---KIA-----488  
Db 4271 DKEGVWTKCNKKTLDLRLVTGLTEGHSYEFVRVAENNAAGVGEPSVVFYRACDALY 4330  
QY 489 -----VMD--R-----IY-----TK---D-----497  
Db 4331 PRGPPSNPKVTDTSRSSVSLAWSKPIYDGGAPVKYGVVVEVKEAADAETWTCTPTTGLQK 4390  
QY 498 -----LK-NA--F-----KKA-----MD-----507

Db 4391 QFTVTKLKENTENYFRICAINSEGVGEPATLPGSVVAQERIEBPPEIELDADLRKVVVLA 4450  
QY 508 -----KG-----A-----RA-I-----M-----514  
Db 4451 SATLRLFTVIKGRPEPEVKWEKAEGILTDRAQIEVTSSTFMTLVIDNVTRFDSGRYNTLE 4510  
QY 515 -----V-----VN-----517  
Db 4511 NNSGKTAFVNVRLDPSAPVNLITREVKKDSVTLSEWEPPLIDGAKITNIVEKRETT 4570  
QY 518 ---TV-----N-----YY-----N-----R524  
Db 4571 RKAYATIINNCTKTTFRIENLOEGCSYFRVLASNEYGIGLPAETTEPVKVPSEPLPPGR 4630  
QY 525 ---D--N-----W-----T---528  
Db 4631 VTLVDVTRNTATIKWEKPESDGGKITGYVVMQTKGSEKWTCTQVKTLEATISGLTAG 4690  
QY 529 -----ELP-----A-----MG-534  
Db 4691 BEYVFRVAAVNEKGRSDPRQLGVPVIAKDIEIKPSVELPFHTFNVKARQLKIDVPFKGR 4750  
QY 535 -----YE-----A-----537  
Db 4751 POATVNRKDGOTLKETTRVNVSSSKVTYLSLKEASKEDVGYELCYVNSAGSITVPIT 4810  
QY 538 -----DE-----G-----540  
Db 4811 IIVLRPGPPGPIRIDEVSCDITISWNPPEYDGGCOISNYIVEKETTSTTWHIVSOAV 4870  
QY 541 --T--K-----S-O-----V--FS-----547  
Db 4871 ARTSIKIVRLTTCGSIYQFRVCAENRYGKSSYSESAVVAEYFPSPGPGTKPVVHATKS 4930  
QY 548 -----I-----SG-----550  
Db 4931 TMLVTWQVPVNDGGSRVIGYHLEYKERSILWSKANKILLIADTQMKVSGLDGLMVEYRV 4990  
QY 551 -----D-----DG-----V554  
Db 4991 YAENIAGIGKSKCEPVARDPCDPPGOPEVTNITRKSIVLSKWSKPHYDGGAKITGYIV 5050  
QY 555 ---KL---W-----N-----558  
Db 5051 ERRELDPGRWLKCNVTNIQETVFEVTELTEDORYEFRVARNAAADSVSPSSSTGPIIVK 5110  
QY 559 -----M-----IN-----P-----D--KK--TE---567  
Db 5111 DDVEPRVMMDVKPRDVIWVKAQVVKINADIAGRPLPVISWAKOGIEIEERARTEIIST 5170  
QY 568 ---VK-----R--N-----SG-----N--K---573  
Db 5171 DNHTLLTVKDCIRRDGTQYVLTKNVAGTRSVAVNCKVLDKPPGAPLEINGLTAEKCS 5230  
QY 574 -----E---DFKD---K-----L-----E-Q-----Y-----583  
Db 5231 LSWGRPOEDGGADI-DYIYVEKRETSHLAWTICEGELQMTCKVTKLLKNGEYIFRVTVG 5289  
QY 584 --Y---P---I--D-----M-----ES-----F---591  
Db 5290 NKYGVEPLESVAIKALDPFTVPSPPTSLEITSVTKESMTLCSWRPESDGGSEISGYIIE 5349  
QY 592 ---NS-----N-KP--N--V-----G-D-E-----K-----E---I---D-605  
Db 5350 RREKNSLRVVRKNKPVYDLRVKSTGLREGCEYEVYRVAENAAAGLSLPSETSPLIARADP 5409  
QY 606 -F-----K--F--AP-----DTD-K-----614  
Db 5410 VFLPSPPKPKIVDSGKTTITIAWVKPLPDGGAPITGYTVYKKSDDTDMKTSIQSLRGT 5469  
QY 615 E---L---Y-----KE-----DI-----IVPA625  
Db 5470 EYTIISGLTTGAEYVFRVSKVNVKGASDPDSDSDPQIAKEREPEEPLFDIDSEMRKTLIVKA 5529



Oy	626	G-S-T-S-----WG-P-----R-----I-----D-----	635
Db	5530	GASFTMTVPFGRGPVNVLMKSPDLDLTRAVIDTDSRTSLTIENANRNSGKYLTIQ	55698
Oy	636	--L--L-L-K-----P-----DV-----SA-----P-----G--KN--I-----	649
Db	5590	NVLSAASLTIVVKVLDTPGPPNITVQDVTESANLSDVBNQDGAFAVKYHLEKREAS	56499
Oy	650	-K--S-T-LN-----VIN---GKSTY-----G---Y 664	
Db	5650	KKAWSVTNNGNRLSYKVTNLQEG-AIYFVRSGENE 5686	
RESULT 11			
ABB22021	ID	ABB22021 standard; Protein; 5701 AA.	
AC	ABB22021;		
XX	23--JAN-2002	(first entry)	
Dt	Protein #4020	encoded by probe for measuring heart cell gene expression.	
De	Human: gene expression; heart; microarray; vascular system;		
XX	cardiovascular disease; hypertension; cardiac arrhythmia;		
KW	congenital heart disease.		
XX	Homo sapiens.		
OS	MO200157274-A2.		
XX	09-AUG-2001.		
PD	30-JAN-2001; 2001WO-US00666.		
PF	04-FEB-2000; 2000US-0180312.		
PR	26-MAY-2000; 2000US-0207456.		
XX	30-JUN-2000; 2000US-0608408.		
PR	03-AUG-2000; 2000US-0632366.		
XX	21-SEP-2000; 2000US-0234687.		
PR	27-SEP-2000; 2000US-0236359.		
XX	04-OCT-2000; 2000GB-0024263.		
PA	(MOLE-) MOLECULAR DYNAMICS INC.		
PI	Penn SG, Hanzel DK, Chen W, Rank DR;		
XX	WPI; 2001-488899/53.		
DR	Single exon nucleic acid probes for analyzing gene expression in human		
XX	hearts -		
PT	Claim 15; SEQ ID No 23791; 530bp; English.		
PS	The present invention relates to single exon nucleic acid probes for		
XX	measuring human gene expression in a sample derived from human heart (see		
CC	ABR21555-ABR21305). The present sequence is a protein encoded by one such		
CC	probe. The probes may be used for predicting, measuring and displaying		
CC	gene expression in samples derived from the human heart via microarrays.		
CC	By measuring gene expression, the probes are useful for predicting,		
CC	diagnosing, grading, staging, monitoring and prognosing diseases of the		
CC	human heart and vascular system e.g. cardiovascular disease,		
CC	hypertension, cardiac arrhythmias and congenital heart disease.		
CC	Note: The sequence data for this patent did not form part of the printed		
CC	specification, but was obtained in electronic format directly from WIPO		
CC	at ftp.wipo.int/pub/published_pct_sequences.		
XX	Sequence 5701 AA;		
XX	Query Match 69.8%; Score 3070.2; DB 22; Length 5701;		
XX	Best Local Similarity 10.3%; Pred. No. 7e-44;		
XX	Matches 578; Conservative 68; Mismatches 14; Indels 4958; Gaps 495;		

QY	1	Y-----	P-----	VV-----	LA-D-----	T-----	S-----	9																																																
Db	73	YNFRISAINDAVG	GP	PAVI	P	VEI	VER	EMAD	FELDA	ELART	LLV	RAG	LIR	EV	PIGR	132																																								
QY	10	-----	-----	-----	S-S-----	E-----	D-----	A-----	-----	-----	-----	-----	-----	-----	14																																									
Db	133	PAPEVT	TKD	N	L	K	N	R	A	N	I	N	T	E	S	FT	111																																							
QY	15	-----	LN-----	I-----	S-----	D-----	-----	K-E-----	-----	-----	-----	-----	-----	-----	21																																									
Db	193	VLDTP	GP	PLN	R	P	D	I	T	K	D	S	V	T	L	H	M	D	J	P	L	D	I	G	S	R	I	N	T	Y	I	V	E	K	E	A	R	K	S	T	A	T	T	K	H	252										
QY	22	--KV-----	-----	-----	-----	-----	-----	K-E-----	-----	-----	-----	-----	-----	-----	28																																									
Db	253	CTYK	T	G	L	S	E	C	E	Y	F	R	V	M	A	N	E	Y	I	G	E	P	T	E	T	T	E	P	V	K	A	S	E	A	P	S	I	N	I	M	D	I	T	K	S	T	V	S	L	312						
QY	29	-----	KH-----	E-N-----	I-----	-----	-----	H-----	-----	-----	-----	-----	-----	34																																										
Db	313	AMP	K	R	H	G	S	K	I	T	G	Y	V	I	E	A	O	R	K	S	D	O	M	T	H	T	T	V	K	G	L	E	C	V	R	N	L	T	E	G	E	Y	T	F	O	V	A	N	S	A	372					
QY	35	--SA-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	37																																										
Db	373	GRS	A	R	E	S	R	P	V	I	V	K	E	Q	T	M	L	P	E	L	D	L	R	G	I	Y	O	K	V	I	A	K	A	G	N	I	N	V	E	L	P	V	L	G	R	P	K	P	T	Y	T	W	K	K	D	432
QY	38	-----	ET-----	-----	-----	S-----	-----	-----	-----	-----	-----	-----	-----	41																																										
Db	433	QIL	K	O	T	O	R	V	N	E	T	A	T	S	T	I	L	I	N	E	C	R	S	D	S	G	P	P	L	A	R	N	I	V	E	G	V	D	I	T	I	O	V	H	D	I	P	G	P	T	492					
QY	42	-----	DF-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	44																																										
Db	493	GPI	K	E	D	E	V	S	D	P	F	T	S	W	D	P	P	E	N	D	G	V	P	I	S	N	V	V	E	M	R	O	T	S	T	T	W	E	L	A	T	T	I	R	T	Y	K	A	T	R	L	552				
QY	45	--E-----	K-----	-----	-----	K-----	-----	TAV-----	-----	-----	-----	-----	-----	50																																										
Db	553	T	T	G	E	Y	O	R	V	A	K	O	N	R	Y	G	V	G	P	G	I	T	S	A	C	I	A	N	T	P	F	K	V	P	G	P	T	O	V	A	T	A	T	K	D	S	M	T	S	M	E	P	L	612		
QY	51	-----	I-----	-----	KE-----	-----	-----	K-----	-----	E-VV-----	-----	-----	-----	57																																										
Db	613	S	D	G	S	P	I	L	G	H	V	E	R	K	E	R	N	G	I	M	O	T	V	S	K	A	L	V	P	G	N	I	F	K	S	G	L	D	G	A	I	E	F	R	V	I	A	E	M	N	A	K	S	K	672	
QY	58	-SK-----	-----	-----	N-----	-----	-----	P-----	-----	V-I-----	D-----	-----	-----	64																																										
Db	673	P	S	K	P	E	P	M	L	A	D	P	I	D	P	G	K	P	V	P	L	N	T	R	H	T	V	T	L	K	N	A	K	P	E	Y	T	G	F	K	I	T	S	Y	I	E	K	R	D	L	P	N	G	R	W	732
QY	65	-----	N--N--T--S-----	-----	-----	N-----	-----	E-----	-----	BA-KIX-75	-----	-----	-----	75																																										
Db	733	L	K	A	N	F	S	N	I	L	E	N	E	F	T	S	G	L	T																																					

QY 114 -E- -S- -GE- -KA- - 119  
Db 1153 NOEYKIRIYAMNSEGLGEPALVPGTPKAEADRMPLPEIELDADLRKVVIRACCTLRFLVP 1212  
QY 120 IK- -E- -L- -SS- -L- -K- - 127  
Db 1213 IKGRPAPEVKWARDHGESLDKASIESTSYLLIIVGNVNRFDGKYILTVENSSEKSAF 1272  
QY 128 -N- -TK- -V- -L- -Y- -T- -Y- - 135  
Db 1273 VNVRLDTPGPPQDLKVAEVTKTSVTLTWDPLDGGKIKNYIVKRESTRKAYSTVAT 1332  
QY 136 -DRI- -F- -N- -G- -S- -A- - 143  
Db 1333 NCHKTSWKVDLQEGCSYFVRVLAENEYIGLPAETABSVKASERPLPPGKITLMDVTRN 1392  
QY 144 -I- -E- -T- -T- - 147  
Db 1393 SVLSWEKPEHGGSRILGYIVEMQTKGSDKWATCATVKVTEATITGLIQGEEYSFRVSA 1452  
QY 148 -P- -D- -N- -L- -D- -K- -I- - 154  
Db 1453 ONEKISDPRQLSVPIAKDLVPPAFKLLFTFTVLAGEDLKVDPVPTIGRPTPAVTHK 1512  
QY 155 -KQ- -I- -IE- - 158  
Db 1513 DNVPLKQTRVNAESTENNSLLTIKDAREDVGHVYVVKLTNSAGEAETLNLNVLDKPGP 1572  
QY 159 -G- -I- -SS- -VE- -R- -A- - 166  
Db 1573 PTGPVKMDEVTAADITLSWGPKYDGGSSINNYIVEKRDSTTTTQWIVSATVARTIKAC 1632  
QY 167 -Q- -K- -V- -Q- -P- -M- -M- -N- - 174  
Db 1633 RLKTCGEQFRIAENRYGKSTYLNSEPTVAQYPEKVPGPPTVVLTSSRDSMEVQWNE 1692  
QY 175 -H- -ARKE- -I- -GVE- - 183  
Db 1693 PISGGSRVIGYHLERKERNLSLWVKLNKTPITQTKFTTGLEGEVEYEFVRVSAENIVGI 1752  
QY 184 -EAI- -D- -Y- -L- - 189  
Db 1753 GKPSKVSECVYARDPCDPGRPEALIVTRNSVTLQWKKPTYDGGSKITGYIVEKKELPEG 1812  
QY 190 -K- -S- -I- -NA- - 194  
Db 1813 RMMKASFTNIIDHTEFVTGLVEDHRYEFVRVIAARNAAGVPSESESTGALTARDEVDPRI 1872  
QY 195 -P- -F- -CK- -N- -F- - 200  
Db 1873 SMDPKYKOTIVVHAGESFKVDADIYGKPIPTIQTQWIKGOELSENTARLEIKSTDFATSLSV 1932  
QY 201 -D- -G- -R- -G- -MVIS- - 208  
Db 1933 KDAVRVDSGNLYLKAKNVAGERSVTNVKVLDRPGPPEPVVISGVTAEKCTLANKPPLQ 1992  
QY 209 -N- -I- -D- -T- -G- - 213  
Db 1993 DGGSDIINIVERRRETSLRVMTVDANVQTLSCVKVTKLEGNEYTFRIMAVNKYGVGEPL 2052  
QY 214 -T- -D- -Y- - 216  
Db 2053 ESEPVVAKNPFVVPAPKAPEVTTVTKOSMIVVWERPASDGGSEILGYVLEKRDKEGIRW 2112  
QY 217 -R- -H- -KA- - 220  
Db 2113 TRCHKELICELRLVTGLIENHDYBFRVSAENAGLSEPSPEYAKACDPYKPGPPNN 2172  
QY 221 -M- - 221  
Db 2173 PKVIDITRSVFLSKPIYDGGCEIQGYIVEKCDVSUGEWMTMPTPTGINKNTINEVEKL 2232  
QY 222 -RI- -D- -D- -D- -AK- -A- -SMR- -F 233

Db 2233 LEKHEYNFRICAINKAGVGEHADVPGPPIIVEKLEAPDIDLDELURKIINIRAGESLRLF 2292  
QY 234 -K- -K- -E- -D- -L- -K- -GT- - 241  
Db 2293 VPIKGRPTPEVKWGVGEIRDAALIIDVTSSFTSLVDNVRNRYDSKGYTLTLENSSGTKS 2352  
QY 242 -N- -D- -KNY- - 245  
Db 2353 AFVTVRVLDTSPSPVNLKVTBITKDSVSITWEPPLDGGSKIKNYIVEKREATRKSAAV 2412  
QY 246 -W- -L- -S- -D- -KI- - 251  
Db 2413 VTNCHKNSWKIDQLQEGCSYFVRVTAENEYIGLPAQTADPIKVAEVPQPGKIIVDDVT 2472  
QY 252 -P- -H- -A- -F- - 255  
Db 2473 RNSVLSWTKPEHDGGSKIIQIVIVEMQAKHSEKSECARVKSLOQAVITNLTOGEBYLFV 2532  
QY 256 -N- -Y- -Y- - 258  
Db 2533 VAVNEKGRSDPRSLAVPIVAKDLVIEPDKPAPSSYSVQVGDLEKIEVPISGRPKPTITW 2592  
QY 259 -N- -GGK- -ITV- -E- - 266  
Db 2593 TKDGLPLKQTRINVTDSLDLTLSIKETHKDDGGQYGITVANVVGOKTASIEIVTLDKP 2652  
QY 267 -KYDD- -G- -RD- - 273  
Db 2653 DPPKGPVKFDDVSAESITLSWNPPLYTGCGQITNYIVQKRDTTITVWDVVSATVARTTLK 2712  
QY 274 -Y- -P- -DP- - 277  
Db 2713 VTKLKTGYOFRIFAENRYGOSFALSDPIVAQYPYKEPGRPFPFATAISKDSMVIQW 2772  
QY 278 -GMH- -I- -A- -GI- -LA- - 287  
Db 2773 HEPVNNGGSPVIGYHLERKERNLSILWTKVNTKIHTDTQFKAQNLBEGIEYEFVRVAENIV 2832  
QY 288 -G- -N- -D- -T- -E- -Q- -D- -I- -K- - 296  
Db 2833 GVGKASKNSECVYARDPCDPPTGPPIVMKREITLQWTKPVYDGGSMITGYIVEKRDLP 2892  
QY 297 -N- -F- -N- -G- -I- -D- -G- -I- -A- - 305  
Db 2893 DGRMMKASFTNVITQFTVSGLTEDQRYEFVRVIAKNAAGAIKSPSDSTGPITAKDEVLP 2952  
QY 306 -P- -NA- - 308  
Db 2953 RISMDPKPRDTIVVNAGETFRLEADVHGKPLPTIENLRGDKIEESARCEIKNTDFKALL 3012  
QY 309 -Q- -I- -F- - 311  
Db 3013 IVKDAIRIDGGYILRASNVAGSKFPVNVKVLDRPPEGPVQVGTVTSEKSLTWSPP 3072  
QY 312 -S- -Y- -K- -M- -Y- - 316  
Db 3073 LQDGGSDISHYVVEKRETSRLAWTVVASEVNTSLUKVTKLLEGNEYVFRIMAVNKYVGE 3132  
QY 317 -SDAG- -SG- - 322  
Db 3133 PLESAPVLMKNPFFVLPGPPKSLLEVNTNIAKDSMTVCWNRPSDGGSEIIGYIVEKDRSGI 3192  
QY 323 -F- -AG- -DETWF- - 330  
Db 3193 RWIKCNKRRTDLRLRVTLGTEDHEYEFRVSAENAAAGVEPSPATVYVYKACDPV-FKPGP 3251  
QY 331 -H- -A- -I- -E- -D- -S- -I 337  
Db 3252 PTNAHIVDTTKNSITLAWGKPIYDGGSEILGYVVEICKADEEEOIIVTPQTGLRVTRFEI 3311  
QY 338 -K- -H- -N- -V- -D- -V- -S- - 345

Db 3312 SKLTHEOEYKIRVCAINLKVGLGEATSVBGTVPKPEDKLEAPELIDSELRKGIIVVRAGSSA 3371  
 QY 346 -S-S--G-FT-----G--T-----G-----L----- 355  
 Db 3372 RIHIFPKGRPEPPEITWSHEBEFTDKVOIEKGVNVTQLSINCDNRDAGKYLKLENSSG 3431  
 QY 356 -V-----G-----E--K--Y-WQ-----A-----I-----R--- 365  
 Db 3432 SKSAFVTVKVLDTGPPONLAVKEVRKDSAPLWMEPPIIDGAKVKNVVIDKRESTRKAY 3491  
 QY 366 A-----L-----R-KA-----GIP-----MV--- 375  
 Db 3492 ANVSSKSKTSFKVENLTEGAIYYFRVAENEFVGVEVETVDAVKAAPSPSPGKVTLT 3551  
 QY 376 -V---AT-----G-----N-----Y- 381  
 Db 3552 DVSQTSASLWMEKEPBHGDGSRVLGVVEMQPKTEKMSIVAESKVCNAVVTGLSSGOEYQ 3611  
 QY 382 -A-T--S-----A-----S--S----- 388  
 Db 3612 FRVKAYNKSGKSDPRVLGVPVIAKDLTIQPSLKLBFNYTSIOAGEDLKIPIVIGRPRPN 3671  
 QY 389 -SW--D--L-----VAN-----N--HLK-----MT----- 401  
 Db 3672 ISWVWDEGPELTKQTRV--NVEETATSTVLIKEGNKDDFGKYTVTATNAGTATENLSIV 3730  
 QY 402 -----D-----TG-----N-VT--R--TAA--H----- 412  
 Db 3731 LEKGPVGVPRPEVSADPFAVISMEPPAYTGCOISNYIVEKDKTTTTHMVASATVAR 3790  
 QY 413 -----E-----D--A-IA-----VAS-AK--- 422  
 Db 3791 TTIKITKLKGTGEVQFRIFAENRYKSAAPLDSKAVIYQYPPKEPPEPTPTVTSISKQOM 3850  
 QY 423 -N-----Q-----T-V--EF-----D-----KV- 431  
 Db 3851 LVQWHEPVDGDTKLIIGYHLEQKEKNSILMWKLNKTPIDTKFKTTGDEGLEVEFKVSA 3910  
 QY 432 -N--G-GB--S--F-----K--Y--- 440  
 Db 3911 ENIVIGIKRPSKVSCEGFARDCDPPRPEAIVITNNVTLMKKKRAYDGSKITGYIEK 3970  
 QY 441 -R-----N--G-----A--F----- 446  
 Db 3971 KDLDGGRMKKASFTNVLETFEFTVSGLVEDQRYERFVIRMAAGNSEPSDSSGATARDE 4030  
 QY 447 -----F--D--K-----SK-----I 452  
 Db 4031 IDAPNASLDPKYKDVIVVHAGETFLVLEADIRGKPIPDVWMSKDGKELETAARMEIKSTI 4090  
 QY 453 -TT-----N-----E--D-----G-T--K-- 460  
 Db 4091 QKTTLVVKDCIRTDGQYILKLSNVGKTSIPIVTKVLDPRGPBGPLKVTGVAEKCYL 4150  
 QY 461 A--P-----S--K-----L--K-----F--V-- 468  
 Db 4151 AMNPLODGCANISHIYIEKETSRLSMTQVSTEQALNKYTKLLPBGNEVIFRMAVANK 4210  
 QY 469 Y-IG-----K--G--O-----D-----Q--D-----LI--- 479  
 Db 4211 YGIGPLESGPVTA CNPKPPGPSTPEVSATIKDSMVTVARPVDDGTEIEGYILEKR 4270  
 QY 480 -G-----L-DLR-----G-----KLA----- 488  
 Db 4271 DKEGRWTKCNKKTTLTLRLRVLTGLTEGHSYEFVAAENAGVGSEPSVFRACALY 4330  
 QY 489 -VMD--R-----Y-----TK-----D----- 497  
 Db 4331 PRGPSPSNKVTDTSSSVSLAMSKPIYDGCAPVKGVVVEKAADDEMTTCTPTGLOGK 4390  
 QY 498 -LK-NA--F-----KKA-----MD----- 507  
 Db 4391 QFTVTKLKENTENYFRICAINSEGVGEPA TLPGSVVAQERIEPPEI ELDADLRKVVVLR 4450

QY 508 -S-S--G-FT-----G--T-----G-----L----- 355  
 Db 4451 SATIRLFVTIKGRPEPEVKEKAEAGILTDRAQIEVTSFTMLVINDVTRPDSGRNLTLE 4510  
 QY 515 -V-----G-----E--K--Y-WQ-----A-----I-----R--- 365  
 Db 4511 NNSGSKTAFVNVVRVLDSPAPVNLTIREVKKDSVTLSEWPEPLIDGAKITNYIVEKRETT 4570  
 QY 518 -TV-----N-----Y-----N----- 524  
 Db 4571 RKAVATTNNCTKTTPRIENLQEGCSYFVRVLSNEXYIGLPAETPEPVKSEBPLPGR 4630  
 QY 525 -D--N-----W-----T----- 528  
 Db 4631 VTLVDVTRNATTIKMEKPEBDSGSKITGVVENMOTKSEKSKTCTQVKTLEATISGLTAG 4690  
 QY 529 -----ELP-----A-----MG- 534  
 Db 4691 EEYFRVAAVNEKGRSDPRLQGVPIARDIEIKPSVELPFHTFNVKAREQLKIDVPFKGR 4750  
 QY 535 -----YE-----A----- 537  
 Db 4751 POATVNRKDGQTLKETTRVNVSSSKTVTSLSIEAKSEKEDVGTVELCVNSAGSITVPIT 4810  
 QY 538 -DE-----G----- 540  
 Db 4811 IIVLDPRGPGRPIRIDEVSCDSTITSMNPPEYDGGCOISNYIVEKKTSTTHHIVSOAV 4870  
 QY 541 -T--K--S--Q-----V--FS----- 547  
 Db 4871 ARTSIKIVRLTSGEYQFRVCAENRYKSSVSESSAVVAEBPFPBPGRPKVVAHTKS 4930  
 QY 548 -N-----I-----SG----- 550  
 Db 4931 TMLVQVQPVNDGSRVIGYHLEKERSSILMSKANKILADTOMKVSGLDEGLMEYERV 4990  
 QY 551 -----D-----DG-----V 554  
 Db 4991 YAENIAGIKCSKSCBPVPAPDCDPPGOPEVNTITRSVSLKMSKPHYDGAKITGYIV 5050  
 QY 555 -KL--W-----N----- 558  
 Db 5051 ERRLEPGRMWLKKNYTNIGETFEVTELTEDQRYERFVIRMAADSVSEPSGTPIVK 5110  
 QY 559 -N-----IN-----P-----D--K--TE--- 567  
 Db 5111 DVEPPRVMDVKFERDVIIVVAGEVLKINADIAGRPLPVISWADGIEIEERARTEIIST 5170  
 QY 568 -VK-----R--N-----N-----K-- 573  
 Db 5171 DNHTLLTVKDCIRRDQGYVLTLMKNVAGTNSVAANKVLDKPPGAPLEITNGLTAEKCS 5230  
 QY 574 -E--DFKD--K--L-----E--Q-----Y 583  
 Db 5231 LSWGRPEQDGAADI-DYIYIEKRETSHLAWTICGELQMSCKYTKLLKNGEYIFRVTGV 5289  
 QY 584 -Y--P-----I--D-----M-----BS-----F-- 591  
 Db 5290 NKYGVEPLESVAIKALDPFTVPSPPTSLBITSTYKESMTLWMSRPEBDGSEISGYIIE 5349  
 QY 592 -NS-----N--KP--N--V-----G-D-E--K-----E--I--D- 605  
 Db 5350 RREKNSLRWVRVNNKFPYDLRVKSTGRREGCEYERYVAENAGLSLPSESPILRAEDP 5409  
 QY 606 -F-----K--F--AP-----DTD--K----- 614  
 Db 5410 VFLPSPSKPIVDSKTTITIAVAVKPLFDGAGIITGYTVEYKKSDDTDWKTSTISQSLRG 5469  
 QY 615 E-----L-----Y-----KE-----DI-----IYPA 625  
 Db 5470 EYITISGLTGAELYVFRVSVNKGASDPSPDSDPOIAKEREERBPFLIDSEMRKTLIVKA 5539

QY 626 G-S-T-S-----WG-P-----R-I-----D----- 635  
Db 5530 GASFTWTFGRGPPVNVLSKPDLDLRTRAVDTTDSRTSLTIENARNDSGKVTLTQ 5589  
QY 636 -L-L--L--K-----P-----DV--SA-----P--G--KN-I----- 649  
Db 5590 NVLSAASLTLLVVKVLDTPGPTNITVQDVTKESAVLSMDVPENDGGAPVKNVHIEKREAS 5649  
QY 650 -K--S-T--LN-----VIN-----GKSTY-----G--Y 664  
Db 5650 KKAWSVTNNCNRLSYKVTNLOEG-AIYFRVSGNEF 5686

## RESULT 12

AAM57448  
ID AAM57448 standard; Protein; 5701 AA.

XX AAM57448;

XX 05-NOV-2001 (first entry)

DE Human brain expressed single exon probe encoded protein SEQ ID NO: 29553.

XX Human; brain expressed exon; gene expression analysis; probe;  
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
KW epilepsy; cancer.

XX Homo sapiens.

XX WO200157275-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00667.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human  
PT brains -

XX Example 4; SEQ ID NO: 29553; 650pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is a protein encoded by one of  
CC the probes of the invention.

XX Sequence 5701 AA;

Query Match 69.8%; Score 3070.2; DB 22; Length 5701;

Best Local Similarity 10.3%; Pred. No. 7e-44;

Matches 578; Conservative 68; Mismatches 14; Indels 4958; Gaps 495;

QY 1 Y-----P-VV--LA-D-----T-----S----- 9

Db 73 YNFRISAINDAGVGPAVIPDVEIVEREMAPDFELDAELRLTLVVRAGLSIRIFVPIKGR 132

QY 10 -----S-S-----E-----D-----A----- 14

Db 133 PAPEVTTKDNLNKNRANIENTESFTLLIIPECNRYDTGKFMVTIENPAGKKSQVNVNR 192  
QY 15 -----LN-----I--S-----D-----K-E----- 21  
Db 193 VLDTPGPVNLNRPDITKDSVTLHWDLPIDGSGRITNYIVEKREATRKSYSATTATKCHK 252  
QY 22 --KV-----AEN-----K-E----- 28  
Db 253 CTYKVTGLSEGEYFFRVMANEYIGIPTETTEPVKASEAPSPDPSLNIMDITKSTVSL 312  
QY 29 -----KH-----E-N--I-----H----- 34  
Db 313 AWPXKHDGSGKITYVIEAQRKSGDQWTHITTVKGLCEVVNRNLTEGEYTFQWMAVNSA 372  
QY 35 --SA-----M----- 37  
Db 373 GRSAPRESRPVIVKEQTMPELDLRGIYQKLVIKAGDNIKVEIPVLRPKPTVTWKXGD 432  
QY 38 -----ET-----S-----Q----- 41  
Db 433 QILKQTVNFETTATSTILNINECVRSDSGPYPLTARNIVGEVGDVITIQVHDIPGPPT 492  
QY 42 -----DF-----K----- 44  
Db 493 GPIKFEVSSDFVTFSWDPDNDGGVPISVYVEMRQDSTTWVELATTIVRTTYKATRL 552  
QY 45 --E-----K-----K-----TAV----- 50  
Db 553 TTGLEYPQVRKAQNRVGVGPGITSACIVANYPFKVPGPPGTQVTTAVTKDSMTISWHEPL 612  
QY 51 -----I-----KE-----K-----E-VV----- 57  
Db 613 SDGSPILGYHVERKERNGILQWTVSKALVPGNIFKSSGLTDGIAFERVIAENMAGSK 672  
QY 58 -SK-----N-----P-----V-----I-----D----- 64  
Db 673 PSKSEPMALDPIPPGKPYPLNITRHTVTLKWKAPETVGGFKITSIVKRDLPNGRW 732  
QY 65 -----N--N--T--S-----N-----E-----E-KIK- 75  
Db 733 LKANFSNILENEFTVSGLTEDAAAEFRVIAKNAAGAIAPPSEPSDAITCRDDVEAPKIV 792  
QY 76 -----I-----E-----E-----N-SNK- 81  
Db 793 DVFKDQTVILKAGEAFRLADVSGRPPTMWSKDGKEGTAKLEIKIADPSTNLVNDK 852  
QY 82 S--Q--GDYT-----D-----S-F----- 90  
Db 853 STRDSGAYTLTATNPGGFAKHIFNVKVLDRPGPEPGLAVTEVTSEKCVLSWFPPLDDG 912  
QY 91 -----V-----N-----K-----N-----T 95  
Db 913 GAKIDHYIVQKRETSRLAWTNVASEVQVTKLVTKLLKGNEYIFRMAVNAKYGVGPLES 972  
QY 96 E-----N-----PK-----K-----E-DK-----V- 104  
Db 973 EPVLAVNYPGPPDPKPNPEVTTITKDSMVVCGHPDSDGGSEIINYIVERDKAGQRMWK 1032  
QY 105 -----V-----Y-----I-AE-----PK----- 111  
Db 1033 CNKKTLDRLRYKVSGLTEGHEVEFRIMAENAGISAPSPFPFYKACDTVFKPGPGNPR 1092  
QY 112 -----D-----K----- 113  
Db 1093 VLDTSRSSISIAWNKPIYDGGSEITGYMVEIALPEDEWQIVTPPAGLKATSYTITGLTE 1152  
QY 114 --E-----S--GE-----KA----- 119  
Db 1153 NOEKIRIYAMNSEGLGEPALVPGTPKAEEDRLMPPPEILDADLRKVTIRACCTLRLFVP 1212  
QY 120 IK-----E-L-----SS-----L-----K----- 127

Db 1213 IKGRPAPEVKWARDHGSGLDKASIESTSSYTLIIIVGNVRPDSGKYILTVENSSGSKSAF 1272  
 Qy 128 -N-----TK-V-----L-----Y-----T-----Y----- 135  
 Db 1273 VNVAVLDTPPGPPDLKVKAEVTKTSVTLTWDPPLLDGSGKIKNIYIVEKRESTRKAYSTVAT 1332  
 Qy 136 -----DRI-----F-----N-----G-----S-----A----- 143  
 Db 1333 NCHKTSWKVQDLQEGCSYFPRVLAENEYIGLPAPETASVKASERPLPGKITLMDVTRN 1392  
 Qy 144 -----I-----E-T-T----- 147  
 Db 1393 SVSLSWEKPEHDGSRILGIYIVEMQTGSDMKATCATVKTEATITGLIOGEEVSFRVSA 1452  
 Qy 148 -----P-----D-----N-----L-----D-K-----I----- 154  
 Db 1453 ONEKGISDPROLSVFIKADLVIPPAFKLFTFTVLAGEDKLDKDVPIGRPTPAVTWK 1512  
 Qy 155 -----KQ-----I-----E----- 158  
 Db 1513 DNVPLKQTRVNAESTENNSLLTIKDACREDVGHVVKLTNSAGAIETLNVILDKRCP 1572  
 Qy 159 -G-----I-----SS-----VE-----R-----A----- 166  
 Db 1573 PTGPKMDEVTAADSTILSMGPPKYDGGSSINNYIVEKDTSTTWQIVSATVARTTIKAC 1632  
 Qy 167 -----Q-----K-----Y-Q-----P-----M-----M-----N----- 174  
 Db 1633 RLKTGCEYQFRIAEENRYGKSTYLNSEPTVAQYEPKVGPPGTPTLSSRDSMEVQNE 1692  
 Qy 175 -----H-ARKE-----I-----GVE----- 183  
 Db 1693 PISDGSNVIGYHLERKERNSLMWKLNKTPIPQTKFTTGLEBEVEIEFRVSAENIVGI 1752  
 Qy 184 -----EAI-----D-----Y-----L----- 189  
 Db 1753 GKPSKSECVAVARDCDPPGRPEALIVTRNSVTLQMKKPYDGSKITGYIVEKKELREG 1812  
 Qy 190 -K-S-----I-NA----- 194  
 Db 1813 RWMKASFTNIIDTHEVTGLVEDHRYEPRVIARNAAGVFSPESEGTGAITARDEVPPRI 1872  
 Qy 195 -P-----F-----GK-----N-----F----- 200  
 Db 1873 SMDEKYKDTIVNHAGESFKVDADYKRPITTIQWIKGQELSTNARLEIKSTDPAISLSV 1932  
 Qy 201 -D-----G-----R-----G-WVIS----- 208  
 Db 1933 KDAVRVDSGNVILAKKNVAGERSVTNVNKKVLDPRPGPEGPVVISGVTAKECTLAMKPEPLQ 1992  
 Qy 209 -N-I-----D-----T-----G----- 213  
 Db 1993 DGGSDIINYIVERRETSRLVWTVVDANVQTLSCVKTLLEGNEYFRIMAVNKYGVGEPL 2052  
 Qy 214 -----T-----D-----Y----- 216  
 Db 2053 ESEPVNAKPPVNDAPKAPREVTYTKDSMIWVERPASDGSSELGLVEKRDKEGIRW 2112  
 Qy 217 -----R-----H-----KA----- 220  
 Db 2113 TRCHKRLIGELRLKRTGLIENHDYEFVSAENAGLSEPPSAVQKACDPIYKRGPPNN 2172  
 Qy 221 -----M----- 221  
 Db 2173 PKVIDITRSSVFLSKSPIYDGCCEIOGYIVEKCDVSGEWTMCTPPTGINKTINIEVKL 2232  
 Qy 222 -----RI-----D-D-D-AK-----A-SMR-F 233  
 Db 2233 LEKHEYNFRICAINKAGVGEHADVGPPIIVEBKLEAPRIDDLERKIIINRAGSGSLP 2292  
 Qy 234 -----K-K-E-----D-----L-----K-----GT- 241  
 Db 2293 VPIKGRPTPEVWKGVDEIRDAIIVDTSSFTSLVLDNVNRVDSGKYTLTLENSSGTKS 2352

Qy 242 -----D-----KNY----- 245  
 Db 2253 AFVTVRVLDTPSPPVNLKVTIEITKDSVITWEPPLLDGGSKIKNIYIVEKREATRKSAAV 2412  
 Qy 246 -----W-L-----S-----D-----KI----- 251  
 Db 2413 VTNCHKNSWKIDQLEGCSYFVRVTAENEXIGLPAQOTADPIKVAEVPQPGKITVDDVT 2472  
 Qy 252 -----P-H-----A-----F- 255  
 Db 2473 RNSVLSWTKEHDGSKIIIOYIVEMQAKHSEKNSCARVKSLOAVITNLTOGEEYLFV 2532  
 Qy 256 -N-----Y----- 258  
 Db 2533 VAVNEKGRSDPRSLAVPIVAKDLVIEPDVKAPESSYSVQGDUKIEVPISGRPKITTW 2592  
 Qy 259 -----N-----GK-ITV-----E----- 266  
 Db 2593 TKDGLPLKQTRIVNTDSDLTTLTIKETHKDDGGGYGIVANVVGOKTASIEIVTLDKP 2652  
 Qy 267 -----KYDD-----G-----RD----- 273  
 Db 2653 DPKGPKVFDVSAESITLSNPNPLYTGGCOITNYIVOKRDTTTTVDVVSATVARTTLK 2712  
 Qy 274 -----Y-----F-----DP----- 277  
 Db 2713 VTKLKTGYQFRIAEENRYGQSFALSDPIVAQYKPEGPPTPATAIKDSMWIOW 2772  
 Qy 278 H-----GNH-----I-----A-----GI-----LA----- 287  
 Db 2773 HEPVNGSPYIGYHLERKERNSLMTKANKTIIHDQFOKQNEBGEIEFRVYAENIV 2832  
 Qy 288 -G-N-----D-T-----E-Q-----D-I-----K----- 296  
 Db 2833 GVGRASKNSECVAVARDCDPPGTPEPIVVKRNEITLQWTKPVYDGSKITGYIVEKRDLP 2892  
 Qy 297 -----N-----F-----N-G-I-----D-G-I-A----- 305  
 Db 2893 DGRMMKASFTNVIETQFTVSGLTEDQRYEPRVIAKNAGALSKPSDSTGPIAKDEVELP 2952  
 Qy 306 -P-----NA----- 308  
 Db 2953 RISMDPKFRDTIVNNAGETPRLFADVHGKPLPTIEMLRGDEIESARCEIKNTDFKALL 3012  
 Qy 309 -----Q-I-----F----- 311  
 Db 3013 IVKDAIRIDGQYILRASNVAGSKSPVNVNKKVLDPRPGPEGPVQVTGTSKSLTWSPP 3072  
 Qy 312 -----S-Y-----K-----M-----Y----- 316  
 Db 3073 LQDGSIDSHYVEKRETSRLAMTVASEVNTSLKVTKLENGEYFRIMAVNKYGVGE 3132  
 Qy 317 -----SDAG-----SG- 322  
 Db 3133 PLESAPVLMKNPVLPPGPKSLEVTNIAKDSMTVCNRPDSGSEIIGYIVEKRDNSGI 3192  
 Qy 323 -----F-----AC-----DETWF----- 330  
 Db 3193 RMIKCNKRITDLRLVTLGLTEHDHEYFRVSAENAGVGEPSPATVYVYKACDPV-FKRG 3251  
 Qy 331 -H-----A-----I-----E-----D-----S-I 337  
 Db 3352 PTNAHIVDTTKNSITLAKMKRPIYDGSSEILGYYVEICKADEEWOIVTPOGLRVLTREPI 3311  
 Qy 338 -K-H-----N-----Y-----D-----VV-----S- 345  
 Db 3312 SKLTHEQYKIRVCAINKVGLGEXTSVPGTVKPEBKLEAPBLDLSLRKIVVIRAGGSA 3371  
 Qy 346 -----VS-S-G-FT-----G-T-----G-----L----- 355  
 Db 3372 RIHIFKGRPTPELTWERSGEFTDKVQIEGVNVYTOISIDNCRNDAGKYILKLENSSG 3431

QY 356 -V--G--E-K--Y-WO--A--J--R-- 365  
Db 3432 SKSAFTVVKVLDTPGPONLAVKEVRKDSAFVWEPPIIDGGAKVMYVVDKRESTRKAY 3491  
QY 366 A--L--R-KA--GIP--MV-- 375  
Db 3492 ANVSSKCKTSFKVENLTCAGAIYFRVMAENEFQGVFVETVDVAKAAEPSPGKVTLT 3551  
QY 376 -V--AT--G--N--Y- 381  
Db 3552 DVSOVSASLMWEKPEHDGSRVLGVVEMOPKGTCKMSIWAESKVCNAVVTGLSSGOEQY 3611  
QY 382 -A-T--S--A--S-S-- 388  
Db 3612 FRVAYNEKSKDPRVLGVPIAKDLTIQPSLKLPPFTYSIQAGEDLKIEIPVIGRPRPN 3671  
QY 389 -SW--D--L--VAN--N--HLK--MT-- 401  
Db 3672 ISHWKGEPLKQTRV-NVEETATSVLHIKEGKNKDDFKYTVTATNSAGTATENLSVIV 3730  
QY 402 -D--D--N-VT--R--TAA-H-- 412  
Db 3731 LEKPGPVGVPRDEVSADFWISWEPAYTGGQISNYIVEKRDTTTTTHMVSAVAR 3790  
QY 413 -E--D--A-IA--VAS-AK-- 422  
Db 3791 TTIKITLKTGTQYQFRIFAENRYGKSAPLDSKAVIOYPPKPEGPPGTPFVTSISKDQM 3850  
QY 423 -N--N--Q--T-V--EF--D--KV-- 431  
Db 3851 LVQWHEPVNDGGTKIYHLEQKEKNSILWKLNKTPIQDTKFKTGLDGLCEYEFKVA 3910  
QY 432 -NI-G-E--S-F--K--Y-- 440  
Db 3911 ENIVGIGKSKVSECFVARDCPPGRPEAIVITRNNTLKKWKPAYDGGSKITGYIVEK 3970  
QY 441 -R--NI--G--A-F-- 446  
Db 3971 KDLPGRMKASFTNVLTEFTVSGLVEDORYEPRVIRNAGNFSPSPSSGAIARDE 4030  
QY 447 -F--D--K--SK--I 452  
Db 4031 IDAPNASLDPKYKDVVHAGETVLEADIRGKPIPDVWMSKDGKELEETAARMEIKSTI 4090  
QY 453 -TT--N--E--D--G-T--K-- 460  
Db 4091 QKTTLVVKDCIRTDGGQYILKLSNVGKTSIPITVKVLDLRGPPGEPGLKVTGVTAEKCYL 4150  
QY 461 A--P--S--K--L--K--F--V-- 468  
Db 4151 AMNPLODGGANISHYIIEKRETSRLSWTQVSTEQALNYKVKLLPGNEYIFRMAVVK 4210  
QY 469 Y-IG--K-G--Q--D--Q--D--L-I-- 479  
Db 4211 YGGEPLSGPVACNPKYKPPGSTPEVSAITKDSMVVTWARVDDGGTEIGYILEKR 4270  
QY 480 -G--L-DLR--G--KIA-- 488  
Db 4271 DKEGVRMTCNKKTLDLRLRVTLGLTEGHSYEFRAAENAGVGEPSSEPVYFACDALY 4330  
QY 489 -VMD--R--Y--TK--D-- 497  
Db 4331 PPGPPSNPKVTDTSRSSVSLAWSKPIYDGGAPVKGVYVEVEKAAADEWTTCTPPTGLOGK 4390  
QY 498 -LK-NA--F--KKA--MD-- 507  
Db 4391 QFTVTKLKENYENFRICAINSEGVPATLPGSVVAGERIEPPEIELDLADLRKVVLRA 4450  
QY 508 -KG--A--RA-I--M-- 514  
Db 4451 SATURLFTVTKRPEVKEKABGILTDRAQIEVTSFTMLVIDNVTRFDSGRYNLTLE 4510  
QY 515 -V--VN-- 517

Db 4511 NNSGSKTAFVNVRLDSPSAPVNLTIREVKKDSVTLSWEPPLIDGGAKITWIVIVEKRETT 4570  
QY 518 -TV--N--Y--N--R 524  
Db 4571 RKAYATITNNCTKTTFRIENLOEGCSYFVRVLASNEYIGLPAETTEPVKVSEPLPPOR 4630  
QY 525 -D--N--W--T-- 528  
Db 4631 VTLVDVTRNTATIKWKPESDGGSKITGVVEMOTKGEKWSCTQVKTLEATISGLTAG 4690  
QY 529 -E--ELP--A--MG- 534  
Db 4691 EEEYFVRAAVNEKGRSDRQLGVPIARDIEIKPSVELPPHTFNVKAREQLKIDVPPFKGR 4750  
QY 535 -YE--A-- 537  
Db 4751 POATVNRKDGQTLKETTRVNVSSKTVTSLISKEASKEDVGTVELCVSNAGSITVPIT 4810  
QY 538 -DE--G-- 540  
Db 4811 IIVLDRPGPPGPIRIDEVSCSITISWNPPEYDGCQISNVIVEKETTSTTWHIVSOAV 4870  
QY 541 -T--K--S--Q--V--FS-- 547  
Db 4871 ARTSIKIVRLTTGSEYQFRVCAENRYGKSSYSESAVVAEYFPSPGPGTKPVVHATKS 4930  
QY 548 -I--SG-- 550  
Db 4931 TMLVTQVPVNDGSGRVIGVHLEYKERSIILWSKANKILIAQTOMKVSGLDEGLMYEYRV 4990  
QY 551 -D--DG-- 554  
Db 4991 YAENIAGIGKSKCEPVPARDCPPQPEVTNITRKSVLKWSKPHYDGGAKITGYIV 5050  
QY 555 -KL--W--N-- 558  
Db 5051 ERRELPGRMKCNVNIQETYFEVTELTEDQRYEFRVFNAAADSSEPSSTGPIIVK 5110  
QY 559 -M--IN--P--D--KK--TE-- 567  
Db 5111 DDVEPPRVMMDVKFRDVIWVKGAVLKNADIAGRPLPVIWAKDGIIEERARTEIIST 5170  
QY 568 -VK--R--N--N--K-- 573  
Db 5171 DNHTLLTVKDCIRDTGOYVLTKNVACTRSVAVNCKVLDKPPGPPAGPLEINGLTAEKCS 5230  
QY 574 -E--DFKD--K--L--E-Q--Y-- 583  
Db 5231 LSWGRPQEDGGADI-DYIVIVEKRETSHLAWTICEGELQMTSCVKTKLLKNEYIFRVTGV 5289  
QY 584 -Y--P--I--D--M--ES--F-- 591  
Db 5290 NKYGVEPLESVAIKALDPFTVPSPPTSLEITSVTKESMTLCSRPESDGGSEISGYIIE 5349  
QY 592 -NS--N-KP--N--V--G-D-E--K--E--I--D-- 605  
Db 5350 RREKNSLRWVRNKKPVYDLRVKSTGLREGCEYFVRVVAENAGLSLSETSPLTRAEDP 5409  
QY 606 -F--K--F--AP--DTD-K-- 614  
Db 5410 VFLSPSPKPIVDGSKTTITIAWVKPLFDGAPITGYTVEYKVSDDTDWKTISIOLRGT 5469  
QY 615 E--L--Y--KE--DI--IVPA 625  
Db 5470 EYTTISGLTTGAEYFVRVKSVMKVGASDPDSDPQIAKEREBEPLFDIDSEMRKTLIVKA 5529  
QY 626 G-S-T-S--WG-P--R--I--D-- 635  
Db 5530 GASFTMTVPFRGRPVFNVLWSKPDTLTRAVVDTTDSRTSLTIENANRNDGSKYTLTQ 5589  
QY 636 -L--L--K--P--DV--SA--P--G--KN--I-- 649

Db 5590 NVLSASITLVVKKVLDTPGPPNTITVQDVTKESAVLSNDVBPENDGAEVKNYHIEKREAS 5649

Qy 650 -K---S-T--LN-----VIN---GKSTY-----G---Y 664

Db 5550 KKAMVSVTNNCNRSLSKYKTNLDEG-AIYFVSGSENER 5686

RESULT 13

AAM69843

ID AAM69843 standard; Protein; 5701 AA.

XX AAM69843;

AC

XX

DT 06-NOV-2001 (first entry)

XX

DE Human bone marrow expressed probe encoded protein SEQ ID NO: 30149.

XX

KM Human; bone marrow expressed exon; gene expression analysis; probe;

KW microarray; cancer; leukemia; lymphoma; myeloma.

XX

OS Homo sapiens.

XX

PN W0200157276-A2.

XX

PD 09-AUG-2001.

XX

PF 30-JAN-2001; 2001WO-US00668.

XX

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX

DR WPI; 2001-488900/53.

XX

PT Human genome-derived single exon nucleic acid probes useful for

XX

PS analyzing gene expression in human bone marrow -

XX

PS Example 4; SEQ ID NO: 30149; 658bp + Sequence Listing; English.

XX

CC The present invention provides a number of single exon nucleic acid

CC probes which are derived from genomic sequences expressed in the human

CC bone marrow. They can be used to measure gene expression in bone marrow

CC samples, which may enable the improved diagnosis and treatment of cancers

CC such as lymphoma, leukemia and myeloma. The present sequence is a

CC protein encoded by one of the probes of the invention.

XX

SQ Sequence 5701 AA;

Query March 69.8%; Score 3070.2; DB 22; Length 5701;

Best Local Similarity 10.3%; Freq. NO. 7e-44;

Matches 578; Conservative 68; Mismatches 14; Indels 4958; Gaps 495;

Qy 1 Y-----P--VV--LA-D-----T-----S-----9

Db 73 YNFRISAINDAGVGEPAVIVPEVEIVEREMADPELDALRLRTLVVRAGLSIRIVPIKGR 132

Qy 10 -----S-S-----E--D-----A-----14

Db 133 PAPEVTWKONINLKNRANIENTESFTLLIPECNRVDTGKFWVTIENPAGKSGFVNVVR 192

Qy 15 -----LN-----I--S-----D-----K-E-----21

Db 193 VLDTPGVLNLRPTDITKDSVTLHMDLPLIDGSRITNYIVEKREARTRKSYSTATTKCHK 252

Qy 22 ---KV-----AEN-----K--E-----28

Db 253 CTYKVTGLSEGCCEFFFRMAENEXGIGETTEPVPVASEAPSPBDSLIMIDITKTVSL 312

Qy 29 -----KH-----E-N--I-----H-----34

Db 313 AMPKPKHDSGSKITGVVIEAQRKSGDQWTHITTVKGLCECVARNLTGESEYTFQVMAVNSA 372

Qy 35 --SA-----M-----37

Db 373 GRSAFRESRPVYKEQTMLELDLRLGIYOKLVIAKAGDNIKVEIPVLGRPKPTWKKGD 432

Qy 38 -----ET-----S-----Q-----41

Db 433 QILKQTORVNFETTATSTILINIECVRSDSGPVPLTARNIVGEVDVITIQVHDI PGPPPT 492

Qy 42 -----DF-----K-----44

Db 493 GPIKFEVSSDPVTFPSMDPPENDGVPISNVYVEMROTDSTWVELATTVIRITYKATRL 552

Qy 45 -----E-----K-----TAV-----50

Db 553 TTGLEIYQFRYKAQRVGVGPGITSACTIVANYPFKVPGPGTPOVTAVTXDSMTISHEPL 612

Qy 51 -----I-----KE-----K-----E--VV-----57

Db 613 SDGSPILGYHVERKENGLIMQTVSKALVPGNIFKSSGLTDGIAYEFPRVIAENMAGKSK 672

Qy 58 -SK-----N-----P-----V-----I-----D-----64

Db 673 PSKSEPMALDPIDPPGKVPPLNTHVTTLKWAKEBYTGFKITSYIVEKRLPNGRW 732

Qy 65 -----N--N--T-S-----N-----E-----EA-KIK-75

Db 733 LKANFSINLENEFTVSLTDDAAYEFPRVIAKNAAGAISSPSEPSDAITCRDDVEAPKIKV 792

Qy 76 -----E-----E-----E-----N-SNK-81

Db 793 DVKFKDVTILKAGEAFRLBADVSGRPPTMWSKDGLEGTAKLEIKIADFSTNLVNDK 852

Qy 82 S--O--GDYT-----D-----S-F-----90

Db 853 STRDSGAYTLTATNPGFPAKHIFNVKVLDRPGPPEPLAVTEVTSEKCVLSWPPPLDDG 912

Qy 91 -----V-----N-----K-----N-----T95

Db 913 GAKIDHYIVQKRETSRLAMTNVASEVOVTKLTKLKMEYIFRMAVNVKVGCEPLES 972

Qy 96 E-----N-----PK-----K-----E-DK-----V-104

Db 973 EPVLAVNPPGPPPPKRPPEVTTITKDSMVVCMGHPDSGSEIINYIVERDKAGORWIK 1032

Qy 105 -----V-----Y--I-AB-----PK-----111

Db 1033 CNKKTLTDLRYKVSGLTEGHEYEFRIMAENAGISAPSPFYKACDVTYFKPPGPNR 1092

Qy 112 -----D-----K-----113

Db 1093 VLDTSRSSISIAMNKPIDGSGEITGYMVEIALPEDEMOI VTPPAGUKATSYTTGLTE 1152

Qy 114 -E-----S--GE-----KA-----119

Db 1153 NQEFKIRIYAMNSGCGEPALVPOTPKAEDRMLEPEIELDADLRKVTIRACTLRFLVFP 1212

Qy 120 IK-----E-L-----SS-----L-----K--127

Db 1213 IKGRPAEVMWARDHGSGLDKASIESTSYTLTLVGNVNRPSDGSKYILTVENSSGSKSAF 1272

Qy 128 -N-----TK--V-----L-----Y-----T--Y-----135

Db 1273 VNVKVLDTPGPPPODLKXKEVYTKTSVTLTWDPPLDGGSKIKNYIVEKRESTRKAYSTVAT 1332

Qy 136 -----DRI-----F-----N--G-----S--A-----143

Db 1333 NCHKTSWKVDQLEBGSYFVRVLAENEYIGLPAETAESVKASBRPLPPGKITLMDVTRN 1392  
Qy 144 -----I-----E-T-T----- 147  
Db 1393 SVSLSEKPEHGGSRILGYIVEMQTKGSKWATCATVKVTEATITGLIOGBEYSFRVSA 1452  
Qy 148 -----P-D-D-N-L-D-K-----I----- 154  
Db 1453 QNEKIGSDPROLSVPVIAKDLVIPAPKLLFNTFTVLAGEDLVDPFGIRPTPAVTWHK 1512  
Qy 155 -----KQ-----IF----- 158  
Db 1513 DNVPLKQTRVNAESTENNSLLTIKDACREDVGHVVVKLTNSAGAEIETLNVILDKBGP 1572  
Qy 159 -G-----I-----SS-----VE-----R-----A- 166  
Db 1573 PTGPVKMDEVADSIITLSWGPYPKDGSSINNYIVEKRDSTTTWQIVSATVARTTIKAC 1632  
Qy 167 -----Q-----K-----V-Q-----P-M-----M-----N- 174  
Db 1633 RLKTGCEYQFRIAENRYGKSTYLNSEPTVAQYFPKVPGPPTVVTLSRDSMEVQWNE 1692  
Qy 175 -----H-ARKE-----I-----GVE----- 183  
Db 1693 PISDGSRVIGYHLERKERNILWKLNKTIPOTKFKTTGLEGEVEYEFVRVSAENIVGI 1752  
Qy 184 -----EAI-----D-----Y-----L----- 189  
Db 1753 GKPSKVSCEYVARDPCOPGRPEALIVTRNSVTLQWKPTYGGSKITGYIVEKKELPEG 1812  
Qy 190 -K-S-----I-NA----- 194  
Db 1813 RWMKASFTNIIDTHEFVTGLVEDHRYEPRVIRNAAGVFPSEPSSTGAIARDEVDPRI 1872  
Qy 195 -P-----F-----GK-----N-----F----- 200  
Db 1873 SMDPKYKDTIVVHAGESFKVDADYIGKPIPTQIWKQDQLSNTARLEIKSTDFAATLSV 1932  
Qy 201 -----D-----G-----R-----G-MVIS----- 208  
Db 1933 KDAVRVDSGNYILKAKNVAGERSVTNVVKVLDROPPEGVPVIGVTAETAKCTLAWKPLQ 1992  
Qy 209 -----N-I-----D-----T-----G----- 213  
Db 1993 DGSDDIINYIVERRETSRLVTVVDANVOTLSCKVTKLLENEYTFRIMAVNKYGVGEPL 2052  
Qy 214 -----T-----D-----Y----- 216  
Db 2053 ESEPVAKNPFVDPAPKAPETVTKDSMIVVWERPASDGSSEILGYVLEKRDKEGIRW 2112  
Qy 217 -----R-----H-----KA----- 220  
Db 2113 TRCHKRLIGELRLVGTGLIENHDEYFRVSAENAGLSBPPSPAYQKACDPIYKPGPNN 2172  
Qy 221 -----M----- 221  
Db 2173 PKVIDITRSSVFLSWSKPIYDGCCEIOGYIVEKCDVSGEWTCTPPTGINKNTNIEVKL 2232  
Qy 222 -----RI-----D-D-D-AK-----A-SMR-F 233  
Db 2233 LEKHEYNFRICAINKAGVGEHADVPGPPIIVEKLEAPDIDLDLELRKIINIRAGGSLRLP 2292  
Qy 234 -----K-K-E-----D-----L-----K-----GT-- 241  
Db 2293 VPIKGRPTPEVKWGKVDGEIRDAALIDVTSSFTSLVDNVNRYDSGKYTLTLENSSGTKS 2352  
Qy 242 -----D-----KNV----- 245  
Db 2353 AFVTVRVLDTPSPPNVNLKVTBITKDSVITWEPPLLDGGSKIKNYIVEKREATRKSAAV 2412  
Qy 246 -----W-----L-----S-----D-----KI----- 251  
Db 2413 VTNCHKNSWKIDQLEGCSSYFVRVTAENEYIGLPAQATADPIKVAEVQPPGKITVDVDT 2472

Qy 252 -----P-H-----A-----F-- 255  
Db 2473 RNSVSLSWTKPEHGGSKIIQYIVEMQAKHSEKARVKSLOAVITNLTOGBEYLFVRV 2532  
Qy 256 -----N-----Y-----Y 258  
Db 2533 VAVNEKGRSDPRSLAVPIVAKDLVIEPDVKPAFSSYSVQVGQDLKIEVPISGRPKPTIW 2592  
Qy 259 -----N-----GK--ITV-----E----- 266  
Db 2593 TKDGLPLKQTRINVTDSLDLTLSIKETHKDDGGQYGITVANVVGQKTASIEIVTLDKP 2652  
Qy 267 -----KYDD-----G-----RD----- 273  
Db 2653 DPPKGPVKFDDVSABSIITLSWNPPLYTCGCQITNIVOKRDTTTTVDVVSATVARTTUK 2712  
Qy 274 -----Y--F--DP----- 277  
Db 2713 VTKLKTGTEYQFRIFAENRYGQSFALES DPIVAQYVPYKEPGPGTFFATAISKDSMVIQW 2772  
Qy 278 H-----GMH-----I-----A-----GI-----LA----- 287  
Db 2773 HEPVNNGGSPVIGYHLERKERNISILWTKVNKTIIHDTQPKAQNLBEGIEYEFVRVAENIV 2832  
Qy 288 -G--N-----D--T-----E-Q-----D--I-----K----- 296  
Db 2833 GVKGASKNSECYVARDPCDPGTPPEPIMVKRNEITLQWTKPVYDGGSMITGVIVEKRDLP 2892  
Qy 297 -----N-----F-----N-G-I--D--G-I-A----- 305  
Db 2893 DGRWMAKFTNVVETQFTVSGLTEDQRYEPRVIAKNAAGAIKSPKSDSTGPIAKDEVELP 2952  
Qy 306 -----P-----NA----- 308  
Db 2953 RISMDPKFRDITVNNAGETFRLEADVHGKPLPTIEWLRGDKIEBSARCEIKNTDFKALL 3012  
Qy 309 -----Q-I-----F----- 311  
Db 3013 IVKDAIRIDGGQYILRASNVAGSKSPVNVVKVLDROPPEGVPVQVGTVTSEKSLTWSPP 3072  
Qy 312 -----S-Y-----K-----M-----Y----- 316  
Db 3073 LODGSDISHYVVEKRETSRLAWTVVASEVNTSLKVTLLLEGNEVPRIMAVNKYGVGE 3132  
Qy 317 -----SDAG-----SG- 322  
Db 3133 PLESAPVLMKNPFLVLPGPSPKLEVTNIAKDSMTVCNRPDSDGSEIIGYIVEKDRSGI 3192  
Qy 323 -----F-----AG-----DETWF----- 330  
Db 3193 RWIKCNKRRTDLRLRVLTGLTEDHEYEFRVSAENAGVGEPSPATVYVYKACDPV-FKPGP 3251  
Qy 331 -----H-----A-----I-----E--D-----S--I 337  
Db 3252 PTNAHIVDTTKNSITLANGKPIYDGSSEILGVVVEICKADEBEWQIVTPQTLGRVTRPL 3311  
Qy 338 -K--H-----N-----V-----D-----VW-----S- 345  
Db 3312 SKLTHEQYKIRVCALNKVGLGEATSPGTVKPEDKLEAPDLDELSELKGVVVRAGGSA 3371  
Qy 346 -----VS-S--G--FT-----G--T-----G-----L----- 355  
Db 3372 RIHPFKGRPTPEITWSREGEFTDKVQIEKGVNVTQLSIDNCDRNDAGKYILKLENSSG 3431  
Qy 356 -----V--C-----E-K--Y-WQ-----A-----I-----R----- 365  
Db 3432 SKSAFVTVKVLDTPPQPONLAVKEVRKDSAFVWEPPIIDGAKVKYNYIDKRESTRKAY 3491  
Qy 366 A-----L-----R-KA-----GIP-----MV----- 375  
Db 3492 ANVSSKCSKTSFKVENLTGEGAIYFVRVMAENFVGVPVETVDVAKAABPPSPPGKVTLT 3551



```

QY 376 -V--AT-----G-----N-----Y- 381
Db 3552 DVSQTSASLWMEKPEHDGSRVLGVVEMQTKGTEKMSIVAESKVCNANVTGLSSGQYQ 3611
QY 382 ---A-T---S-----A-----S-S----- 388
Db 3612 PRVAAYNEKGSDBRVLGVPVIAKDLTIQPSLKLPFNTYSIOAGBDLKEIPVIGRPRPN 3671
QY 389 -SW--D--L-----VAN-----N-HLK-----MT----- 401
Db 3672 ISWVKDEPLKQOTTRV-NVEETATSTVLHIKEGNKDFGKXTVTATNSAGATENLSYIV 3730
QY 402 -----D-----TG-----N-VT--R--TAA-H----- 412
Db 3731 LEKGPVPVPRFPEVSADFVVISMEPRAYTGGOISYIVEKDDTTTTHMMSATVAR 3790
QY 413 -----E-----D-A-IA-----VAS-AK-- 422
Db 3791 TTIKITLKTGTQYOFRIFAENRYGSAPLDSKAVIYQYPEKEPQPGTPTVTSISKQOM 3850
QY 423 -----N-----Q-----T-V--EF-----D-----KV-- 431
Db 3851 LVQHEPVDGCTKIIGYHLEQKEKNSILMWKLNKTPIQDTKFKTTGLDEGLEVEFKYSA 3910
QY 432 -NI-G-GE--S--F-----K--Y----- 440
Db 3911 ENIGIGKPSKVSSEGFVARDCDPRGREALVITRNNTLKMKKKRAYDGSKITGYIVK 3970
QY 441 -----R-----NI-----G-----A-F----- 446
Db 3971 KDLPRGWMKASFTNVLETFEFTVSGLVEDQRYEFVIARNAAGNSEPSDSSGAITARDE 4030
QY 447 -----F--D--K-----SK-----I 452
Db 4031 IDAPNASLDPRYKQVIVVHAGETPVLLEADIRKPIPDVVMWSKDGKELETAARMEIKSTI 4090
QY 453 -TT-----N-----E-D-----G-T--K-- 460
Db 4091 QKTLVVKDCIRTDGQYIILKSNVGTGKSIPIYKVLDRGPRPBGPLKGTVAEKYVL 4150
QY 461 A--P-----S--K-----L--K-----F--V-- 468
Db 4151 AMNPLODGGANISHYIIEKRETSRLSWTQVSTEVOALNYKVTKLLPGENEYIFRMAYNKK 4210
QY 469 Y-IG-----K--G--Q-----D-----Q--D-----LI----- 479
Db 4211 YGIEPLESGPVTAQNPYKPPGPSTPEVSAITKDSMNVTVARVVDGTEIEGYILEKR 4270
QY 480 ---G-----L-DLR-----G-----KIA----- 488
Db 4271 DKEGVRMTCKCNKKTLLDLRLKLVGTGLTEGHSYEFVVAENNAAGVGPSEPSVFYRACALY 4330
QY 489 -----VMD-R-----IY-----TK--D----- 497
Db 4331 PRGPSPNPKVDTDRSSSVSLAMSKPIYDGAIPVKGVVEVEKAADDEMTCTPPTGLQGX 4390
QY 498 ---LK-NA--F-----KKA-----MD----- 507
Db 4391 QFTVTKLKENTENFRICAINSEGVEGEPATLPGSVVAGRIEPPPEIIDLARKVVVLA 4450
QY 508 -----KG-----A-----RA-I-----N----- 514
Db 4451 SATLRPLVTIKGREPEVMEKAGSILTDRAQIEVTSFTMLVLDNTRFDSGRYNLTLE 4510
QY 515 -----Y-----VN----- 517
Db 4511 NNSGSKTAFVNVVRLDPSAPVNLITREVKDSVTLSEPPPLIDGAKITNYIVEKRETT 4570
QY 518 ---TV-----N-----YV-----N-----R 524
Db 4571 RKAYAITNNCTKTTFRIENLOEGGSYFRVLASNEYGICLPAETTEPVKVEBPPLPGR 4630
QY 525 ---D--N-----W-----T-- 528

```

```

Db 4631 VTLVDVTRNTATIKMEKPESDGSKITGYVEMQTKSEKSTCTQYVLTLEATISGLTAG 4690
QY 529 -----ELP-----A-----MG- 534
Db 4691 EEYVFRVAANENKGRSDPRQLGVPVIARDIEIKPSVELPFHTFNVKAREOLKIDVPFKGR 4750
QY 535 -----YE-----A----- 537
Db 4751 PQATVMKRCQGLKETTRVNVSSKTVTSLSISEAKEDVGYELCVNSAGSITVPIT 4810
QY 538 -----DE-----G----- 540
Db 4811 IIVLDRPQPGPIRIDEVSCOSITISMNPEYDGCQISNYIVEKETTSTTHIVSQAV 4870
QY 541 -T--K-----S-Q-----V--FG----- 547
Db 4871 ARTSIKIVRLTGTSEYQFRVCAENRYGKSSYSESSAVVAEYFPSPPGPTPKVNHATKS 4930
QY 548 -----I-----SG----- 550
Db 4931 TMLVTQVPVNDGSRVIGYHLEKERSIIMSANKILLADTOMKVSGLDEGLMEYRV 4990
QY 551 -----D-----DG-----V 554
Db 4991 YAENIAGIKKSKSCBPVPARDCDPPGOPEVTNITRKSVSLKMSKPHYDGAKITGYIV 5050
QY 555 -KL--W-----N----- 558
Db 5051 ERRELPGRWMLKCNVTNIQETFEVTELTEDQRYEFVFAFNADSVSEPSSTGPIIVK 5110
QY 559 -----M-----IN-----P-----D--KX--TE----- 567
Db 5111 DDVEPRVMDVKFRDVIIVKAGVLKINADIAGRPLVISMWKGLEIEBRARTEIIST 5170
QY 568 -----VK-----R--N-----N--K-- 573
Db 5171 DNHTLLTVKOCIRRDQGYVTLKNVAGTBSVAANCYLDKPRGPAGLEINGLTAERCS 5230
QY 574 -----E--DFKD--K--L-----E-Q-----Y----- 583
Db 5231 LSMGRPOEDGADL-DYIVIEKRETSRLAWTICEGELQMTSCVKTKLKNGEYIFRVTVG 5289
QY 584 -Y--P-----I--D-----M-----ES-----F-- 591
Db 5290 NKYGVGPELSVALKALDPTVPSPPTSLEITSYTKESMTLQWSRPSDGSSEISGYIIE 5349
QY 592 ---NS-----N-KP--N--V-----G-D-E--K-----E--I--D- 605
Db 5350 RREKNSLRWVRVKKPYYDLRVKSTGLREGCEYERYAENMAGLSLPSEISPLIRADP 5409
QY 606 -F-----K--F--AD-----DTD-K----- 614
Db 5410 VFLPSPSPSKPIVDSGKTTITIAWVKPLFDGAPITGYVEYKKSDDTDMKTSIOSLRGT 5469
QY 615 E--L--L--Y-----Y-----KE-----DI-----IYPA 625
Db 5470 EYITSGLTGAEVYFRVKSVMKVGASDPSDSDPOIAKEREEPLFDISEMRKTLIVKA 5529
QY 626 G-S-T-S-----MG-P-----R--I-----D----- 635
Db 5530 GASFTMTVPFRGRVPRVNVLSKSPDTLRTARAYVDTTOSRFSLTLENNRNRDNGSKYTLTIO 5589
QY 636 -L--L--L--K-----P-----DV--SA-----P--G--KN--I----- 649
Db 5590 NVLSAASLTIVVKVLDPCGPPTNITVQDYTKESAVLSMDVENDGAPVKVNYHIEKEBAS 5649
QY 650 -K--S--T--LN-----VIN--GKSTY-----G--Y 664
Db 5650 KKAWSVTNNCNRLSYKVTNLQEG-ALIVFRVSGENEF 5686

```

RESULT 14

AA05328  
ID AA05328 standard; Protein; 5701 AA.  
AC  
XX  
KW AA05328;  
XX  
DT 09-OCT-2001 (first entry)  
XX  
DE Peptide #4010 encoded by probe for measuring breast gene expression.  
XX  
KW Probe; human; breast disease; breast cancer; development disorder;  
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.  
XX  
OS Homo sapiens.  
XX  
PN WO200157270-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 29-JAN-2001; 2001WO-US00661.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-476286/51.  
XX  
PT Novel single exon nucleic acid probe used to measuring gene expression  
PT in a human breast  
XX  
PS Claim 27; SEQ ID No 14069; 322pp; English.  
XX  
CC The present invention relates to novel single exon nucleic acid probes  
CC (see AA10010-AA10067). The present sequence is a peptide encoded by one  
CC such probe. The probes are useful for measuring human gene expression in  
CC a human breast sample, where the probe hybridizes at high stringency to a  
CC nucleic acid expressed in the human breast. The probes are useful for  
CC predicting, diagnosing, grading, staging, monitoring and prognosis  
CC diseases of the human breast, particularly those diseases with polygenic  
CC aetiology. The diseases include: breast cancer, disorders of development,  
CC inflammatory diseases of the breast, fibrocystic changes, proliferative  
CC breast disease and non-carcinoma tumours.  
CC Note: the sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 5701 AA;

Query Match 69.8%; Score 3070.2; DB 22; Length 5701;  
Best Local Similarity 10.3%; Pred. No. 7e-44;  
Matches 578; Conservative 68; Mismatches 14; Indels 4958; Gaps 495;  
1 Y-----P-VV---LA-D-----T-----S-----9  
73 YNFRISAINDAGVGPVDPDVEIVEREMAPDFELDAELRLTLVVRAGLSIRIFVPIKGR 132  
10 -----S-S-----E-D-----A-----14  
133 PAPEVTWKDNINLKNRANIENTESFTLLIIPCNRYDTGKFVMTIENPAGKSGFVNVR 192  
15 -----LN-----I-S-----D-----K-E-----21  
193 VLDTPGVNLRLPTDITKDSVTLHWDLPLIDGSRITNYIVEKREATRKSYSPTATKCHK 252  
22 ---KV-----AEN-----K-E-----28

Db 253 CTYKVTGLSEGEVFFRVMAENEYIGIGETPTETTEPDKASEAPSPDPSLNIMDITKSTVSL 312  
QY 29 -----KH-----E-N---I-----H-----34  
Db 313 AMPKPKHGGSKITGYVIEAQRKGSQDQWTHITTVKGLECVVRNLTGEBEYTFQVMVNSA 372  
QY 35 --SA-----M-----N-----37  
Db 373 GRSAPRESRPVIVKEQTMPLPELDLRGIYQKLVIAKAGDNIKVEIPVLGRPKPTVTKKGD 432  
QY 38 -----ET-----S-----Q-----41  
Db 433 QILKQTORVNFETTATSTILNINECVRSDSGYPPLTARNIVGEVGDVITIOVHDIPGPT 492  
QY 42 -----DF-----K-----44  
Db 493 GPIKFEVSSDFVTFWMDPPENDGGVPISNVYVEMRQDSTTWLVELATTVINTTYKATRL 552  
QY 45 ---E---K-----K-----TAV-----50  
Db 553 TTGLEIYQFRVKAQNRHYGVGPGITSACIVANYPFKVPGPGTPQVTAVTKDSMTISWHEPL 612  
QY 51 -----I-----KE-----K-----E-VV-----57  
Db 613 SDGSPILGYHVERKERNGILWQTVSKALVPGNIFKSGSLTDCIAYEPRVIAENMAGSK 672  
QY 58 -SK-----N-----P-----V---I---D-----64  
Db 673 PSKSEPMLALDPIPPGKVPPLNTRHTVTLKMAKPEYTGFKITSYIVKRDLPNGRM 732  
QY 65 ---N---N-T-S-----N-----E-----EA-KIK- 75  
Db 733 LKANFSNILENEFTVSGLTEDAAEYFRVIAKAAAGAIAPPSEPSDAITCRDDVEAPKIKV 792  
QY 76 -----N-----E-----E-----N-SNK- 81  
Db 793 DVKFKDVTILKAGEAFLEADVSGRPPPTMWSKDGKEGTAKLEIKIADFTSLVNKD 852  
QY 82 S-Q-GDYT-----D-----S-F-----90  
Db 853 STRRDSGAYTLTATNPGGFAKHIFNVKVLDRPPEGPEGLAVTEVTSEKCVLSWFFPLDDG 912  
QY 91 -----V-----N-----K-----N-----T 95  
Db 913 GAKIDHYIVQKRETSRLAWTNVASEVQVTKLVTKLKGNEYIFRMAVKNKYGVSEPLES 972  
QY 96 E-----N-----PK-----K-----E-DK-----V- 104  
Db 973 EPLAVNYPGPPDPKPKNEVTTITKDSMVVCGHPDSDGSGSEIINVIVERRDKAGQRMWK 1032  
QY 105 -----V-----Y---I-AB-----FK-----111  
Db 1033 CNKKTLLDLRYKVSGLTEGHEGHEYFRMAENAAAGISAPSTSPFYKACDTVFKPGPNR 1092  
QY 112 -----D-----K-----113  
Db 1093 VLDTSRSSISIAWNKPIYDGGSEITGYWVEIALPEDEWQIVTPPAGLAKATSYTTGLTE 1152  
QY 114 --E---S---GE-----KA-----119  
Db 1153 NOEKIRIYAMNSEGLGEPALVPGTPKAEEDRLPPEIELDADLRKVVTIRACCTLRFPV 1212  
QY 120 IK-----E-L-----SS-----L-----K-----127  
Db 1213 IKGRPAPEVKWADHGESLDSKASIESTSSYTLIIIGNVNRNFDGKGYLTVENSSGSKSAP 1272  
QY 128 -N-----TK--V-----L-----Y-----T---Y-----135  
Db 1273 VNVRLDTPGPPQDLKVKVEVTKTSVTLWDPPLDGGSKIKNYIVEKRESTRKAYSTVAT 1332  
QY 136 -----DRI-----F-----N-----G-----S--A-----143  
Db 1333 NCHKTSWKVQLQEGCSYVFRVLAENEYIGLPAETAESVKASERPLPPGKITLMDVTRN 1392

QY 144 -----I-----E-T-T----- 147  
 Db 1393 SVSLSWEKREHDGSGRLIGYIVEMQTKSGDKMCAATVKTVEATITGLIOGEEYSFRVA 1452  
 QY 148 -----P-----D-----N-----L-----D-K-----I----- 154  
 Db 1453 QNEKGISDPRLQSLVPIAKDLVIPPAPKLTLENTFTVLAGEDLKYDVPIGRPTAVTWK 1512  
 QY 155 -----KQ-----TE----- 158  
 Db 1513 DNVPKQOTTRVNAESTENNSLTTIKDAGREDVGHVVKLTNSAGEALETLNVILDKRGP 1572  
 QY 159 -----G-----I-----SS-----VE-----R-----A----- 166  
 Db 1573 PTGVPVKMDEVATADSLTSLMKSPKXDGSSINNYIVEKDDSTTTMQVSAIVARTTIKAC 1632  
 QY 167 -----Q-----K-----V-Q-----P-----M-----N----- 174  
 Db 1633 RLKTGCEYQFRIAEENRYGKSTYLNSEPTVAQYPPKVGPGPTPVVTLSSRDSMEVQNE 1692  
 QY 175 -----H-ARKE-----I-----GVE----- 183  
 Db 1693 PISDGSRVIGYHLERKERNSILWVKLNKTPIPQTKFTTGLEGEVEYEFVSAENIVGI 1752  
 QY 184 -----BAI-----D-----Y-----L----- 189  
 Db 1753 GKPSKVSECVYARDCDPPGPEALIVTRNSVTLOMKKPTYDGSKITGYIVEKKELEP 1812  
 QY 190 -----K-S-----I-----NA----- 194  
 Db 1813 RMMKASFNTNIDTFEVTGLVEDHREYFVIAARNAGVFSESESTGAITARDEVDPRI 1872  
 QY 195 -----F-----GK-----N-----F----- 200  
 Db 1873 SMDPKYKDTIIVNAGSEFKVDADIYKPIPTIOWIKGQELSNARLEIKSTDATSLSV 1932  
 QY 201 -----D-----G-----R-----G-NVIS----- 208  
 Db 1933 KDAVRVDSGNITLAKAVNAGERSVTNVKVLDRPQPPGPPVIGCVTAKEKCTLAMKPPLO 1992  
 QY 209 -----N-I-----D-----T-----G----- 213  
 Db 1993 DGSDDIINYIERETSRLVWTVVDANVQTLSCVTKLLEGNEXTFRIMAVNKKYGVPL 2052  
 QY 214 -----T-----D-----Y----- 216  
 Db 2053 ESEBVAKNPFVVDAPKAPETVTTKDSMIVVWERPASDGSSEILGYVLEKRDKEGIRW 2112  
 QY 217 -----R-----H-----KA----- 220  
 Db 2113 TRCHKRLIGELRLKRVLTGLIENHDYEFVSAENAGLSEPPSPSAVQKACDPIYKRPNN 2172  
 QY 221 -----M----- 221  
 Db 2173 PKVIDITRSSVFLMSKRIYDGCCEIOGYIVEKCDVSGEMTCTPPTGINKTINEVEKL 2232  
 QY 222 -----R-----D-D-D-----AK-----A-SMR-F 233  
 Db 2233 LEKHEVYFRICAINKAGVGEHADVPPIIVEKLEAPDIDLLELRKIIINIRAGSGLLF 2292  
 QY 234 -----K-K-E-----D-----L-----K-----GT----- 241  
 Db 2293 VPIGRPTPEVVKWGVGEIDAAIIDVTSFTSLVDNNVRYSCKYTLTLENSSGTS 2352  
 QY 242 -----KX----- 245  
 Db 2353 AFVTVRVLDTPSPVNLKVTETKDSVSIWEPLDGSKITKIYIVEKREATRKSAAV 2412  
 QY 246 -----W-----L-----S-----D-----KI----- 251  
 Db 2413 VTNCHKSMWKIDLOEGCSYFVRVTAENEYIGLPAQTADEPIKVAEVPQPPGKITVDVT 2472

QY 252 -----P-H-----A-----F----- 255  
 Db 2473 RNSVLSWTKREHDGSGKIIQYIVEMQAKHSEKMSSECAKRVKSLQAVITNLTOGEEYLFV 2532  
 QY 256 -----N-----Y----- 258  
 Db 2533 VAVNEKGRSDPRSIAVPIVAKDLVIEBPDVPAFSSYVQVQDLKIEVPIGRPKPTITW 2592  
 QY 259 -----N-----GGK-ITV-----E----- 266  
 Db 2593 TKDGLPKQOTTRINVTDSLDTLTLISKEITHKDDGQYGITVAVNVGQKTASIEIVTLDKP 2652  
 QY 267 -----KYDD-----G-----RD----- 273  
 Db 2653 DPPKGPVKFDDVSAESITLSMNPPLYTGCGQITNYIVQKDDTTTVDVVSATVARTTLK 2712  
 QY 274 -----Y-----F-----DP----- 277  
 Db 2713 VTKLTGTEYQFRIAEENRYGQSFALSDPIVAQYPKERGPPGTPPATAISKDSMWIOW 2772  
 QY 278 H-----GMH-----I-----A-----GI-----LA----- 287  
 Db 2773 HEPVNGSGPVIGYHLERKERNSILWTKVNKTIIHDTQFAQNLEEGIEYEFVYAEINIV 2832  
 QY 288 -----G-N-----D-T-----E-O-----D-I-----K----- 296  
 Db 2833 GVGRASKNSECVYARDCDPPGTPPEPIWVKRNETILOMTKPVYDGSMTICYIVEKDL 2892  
 QY 297 -----N-----F-----N-G-I-----D-G-I-A----- 305  
 Db 2893 DGRMMKASFNTNIDTFEVTGLVEDQYEFVRIAKNAAGAISKPSDSTGPITAKDEVELP 2952  
 QY 306 -----P-----NA----- 308  
 Db 2953 RISMDPKFRDTIIVNAGETFRLEADVHGKPLPTIEMLRGKLEESARCEIKNTDFKALL 3012  
 QY 309 -----O-I-----F----- 311  
 Db 3013 IVKDAIRIDGQYILRASNVAGSKSFVNVKVLDRPQPPGPPVQVGTSEKSLTWSPP 3072  
 QY 312 -----S-Y-----K-----M-----Y----- 316  
 Db 3073 LQDGSDDISHVVEKRETSRLAWTVASEVVTNLSKVTKLLEGNEXTFRIMAVNKKYGV 3132  
 QY 317 -----SDAG-----SG----- 322  
 Db 3133 PLESAPVLMKNPFVLPBPSPKSLVETNIAKDSMTVCNRPSPDGSSEIIGYIVEKDRSGI 3192  
 QY 323 -----F-----AG-----DETMF----- 330  
 Db 3193 RWICNKRRIITDLRLKRVLTGLTEHDEYEFVSAENAGVGEPSPATVYKACDPV-FKRG 3251  
 QY 331 -----H-----A-----I-----E-----D-----S-I 337  
 Db 3252 PTNAHIVDTTKNSITTLAMKRIYDGSSEILGYVEICKABEEMQIVTPQTLKVTFRFI 3311  
 QY 338 K-----H-----N-----V-----D-----VV-----S----- 345  
 Db 3312 SKLTHQBYXIRVCAKLVKVGISEATSVPGTYKPEDKLEBELDLDSELRKGIIVRAGSA 3371  
 QY 346 -----VS-S-----G-FT-----G-T-----G-----L----- 355  
 Db 3372 RIHIFPKGRPTPEITWSREGEFTDKQIEKGVNYTOLSIDCNRDAGKIYILKLENSSG 3431  
 QY 356 -----V-----G-----E-K-----Y-WO-----A-----I-----R----- 365  
 Db 3432 SKSAFVTVKVLDTGPQPNLAVKEVRKDSAFVWEPPIIDGSAVKVYVIDKRESTRKAY 3491  
 QY 366 A-----L-----R-KA-----GTP-----MV----- 375  
 Db 3492 ANVSSKCSKTSFKVENLTEGAIYFVRVMAENEFQVGVETVDAVKAABPPSPGKVTLT 3551  
 QY 376 V-----AT-----G-----N-----Y----- 381



ID ABG23329 standard; Protein; 6619 AA.  
 AC ABG23329;  
 XX 18-FEB-2002 (first entry)  
 DT Novel human diagnostic protein #23320.  
 DE  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KM food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX MO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 XX 30-MAR-2001; 2001WO-US08631.  
 XX  
 XX 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX (HYSE-) HYSEQ INC.  
 PA  
 PI Drmanac RT, Liu C, Tang YT;  
 DR WPI: 2001-639362/73.  
 DR N-PSDB; AAS87516.  
 XX  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity  
 PS  
 PS Claim 20; SEQ ID No 53688; 103bp; English.  
 XX  
 XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridization probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue; as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WFO  
 CC at ftp.wipo.int/pub/published\_pcc\_sequences.  
 XX  
 SQ Sequence 6619 AA;  
 Query Match 69.5%; Score 3058.8; DB 22; Length 6619;  
 Best Local Similarity 9.1%; Pred.No.2.8e-43;  
 Matches 586; Conservative 63; Mismatches 12; Indels 5802; Gaps 489;  
 QY 1 Y-----P-----VV--LA-----D-----TS-----S----- 10  
 Db 30 FITSPFGKSAIVQIAEDFGKLIDHRLSTSAEDLSGLPRFKENGRAPAPDELFP 89  
 QY 11 -----S-----E-----D-----AL-----N----- 16  
 Db 90 LEGDAVPGFNGWLLFLDEFNSARKEIQAAAYKVLIDMTGQKLLPYVAIVCAGNKATD 149  
 QY 17 -I-----S-----D-----KEKVA-----E-----NK-----EK-H- 30

Db 150 RAITNNIWTAMQSRILHLEMETDFDVFMEVDVAIIPKMDERVIAPLANPNKLNDFEPHQ 209  
 QY 31 ----- 30  
 Db 210 EKTFCSTRTWEFNKIVSILPPGPINSEMTLLAGTITSGVATSFVQFOVYNNVSLNE 269  
 QY 31 -----E--N-----I-----HS-----A--M-----ET--S 40  
 Db 270 ILDNPKARMPEDNNLMAVVTSLINNTDENHSHKIFDYERMPPTFKVLYRSRKTDM 329  
 QY 41 -----Q-----D-----F--K-----E--K-----K-----TA----- 49  
 Db 330 MTDANQLTREVDRQCAKAFMGCKTAAFPQSLCSLKFRRWVKQCCDPAQDTGHELENNPDWE 389  
 QY 50 -----VTK-E-----KEV-----V-----S-----KV-- 60  
 Db 390 SLTPESRYTVMHLEIMHVLHSVRMGSRDPEVNNVACDIYINNQLIDGYFIGIENCW 449  
 QY 61 -P-----V-----I-----D-----D-----NN----- 66  
 Db 450 KDPKYAGMAEQIYDILMSKQIRPPKASGAFGTSGGDMKPSTSKASQVNIWNVBRAM 509  
 QY 67 -----T-----S-----N----- 69  
 Db 510 QQKLSGSLPGVLRGKMEVITQFLKRPVPMQVLLERPFNDLOETTVYSWQRNRYPMY 569  
 QY 70 -----E-----E-A----- 72  
 Db 570 LPSPMDDRLEHLAFLDTSGSIXPKDALFSSSEVAVYKEKQPOKMTLVQFTTEIVDE 629  
 QY 73 -----K-----IK-----E-----E----- 77  
 Db 630 QVIEDGKFTDVTIKGGGTSIVPREWIIHKPTAAIIFTDLEVRMEPLRFDI PVIWV 689  
 QY 78 --N--S-----N--K--S-----Q-----G-----D-- 85  
 Db 690 VIRPAGQVPRGKGMVNSKSSGETTFLGAGVDIRIKQVTVRRFLGLPICKINDEG 749  
 QY 86 -Y-----T-----D-----S--F----- 90  
 Db 750 KTKMKFGFTLASAMERFNVAGLVGHDXSTWARTLSVFNVIIEPGMGVLTLELVY 809  
 QY 91 -----VN--K-----N-----TEN----- 97  
 Db 810 HGRKKLHIPAGSGIAOVNGQKVTVARNAVYTESGKGVITFTNGDSIDTOGYDSIR 869  
 QY 98 -----PK--K--ED-----K-----VW----- 105  
 Db 870 RNVAKALDGAQDIKPKTKYQKTSBDGLANOKAIKANSALVOGKODAVALLVVEQI 929  
 QY 106 -----Y-----I----- 107  
 Db 930 LVALTQAQAKQKQETCTKATGALVWLVMDLADNFEMSMKMGISCVLAREGOELL 989  
 QY 108 -----AEFK-----D-----KE-- 114  
 Db 990 STLPKAEFKLAVTGAHKIVLGSIRTSLTSSYSRLSMINAKILADSITLSDGRVLTMLV 1049  
 QY 115 -----S-----G--EK--A----- 119  
 Db 1050 YPRFIHSEFMTLRALNNASSRAIPTQRLIEOVRENPWPVHWGKQKQASBEELTPM 1109  
 QY 120 -IK--EL-----SS-----L-----K-----N----- 128  
 Db 1110 EIKDAEFLIENVAAGAAVVAEQLRGQVNHQIVRILEPFTHIRVVVTATQWNNPFGRLI 1169  
 QY 129 -----TK-----VL-----Y--T-----YD-----RI----- 138  
 Db 1170 HPDAQPEICELATYMKQYQDSKPRTLTLAGEMHLPTITNSDYLDAYDYCKKHQIRTRDEPS 1229  
 QY 139 -----FNG--SAIE-----T-----TP-- 148

Db 1230 SDEINCLLVKISAAACARASYNFKGTPSTIEEDLRLFSKLVESKPVHASPTEHQATPOH 1289  
QY 149 -----D-----N-----LD-----K--IK-----Q----- 156  
Db 1290 VMSKGLNEPTDPTWENGISHMDRNLRLYSGGHMAKPRFKRDKXPVMSAAGQSHKGD 1349  
QY 157 I--E-----GI-----S-----S-----V-----E--RA-----OK----- 168  
Db 1350 IYVECLQYDGVHAHDTRATELUSKERQLWLSPNKTMIOADLAEVORMLDEWSEDOFIDG 1409  
QY 169 -----V-Q-----P-----N-----M----- 172  
Db 1410 NAEAITDDMASDMVFLQAMAEKPDSTELTTSATLGAQDAMAFGISSEDPAPFQJILSANL 1469  
QY 173 -----M-----N-----HA-----R-----K--E----- 179  
Db 1470 YSNQKLAMVRETLCSNWDHIEGNTDIPHIHTIDKDGFLIFRDYGSIGPKDKIOEVYGV 1529  
QY 180 -----I----- 181  
Db 1530 YGASTKXANTAGTGGFLGCKSPMAYTDSFTVSMCDGKTVYHIAKSSVETNGKPGIIP 1589  
QY 182 --V--EEA-----ID-Y--L-----KSI-----NA-----P----- 195  
Db 1590 IVSVTEESGLEVKIQLQSEDIIDSYIRAILVHGEIKCIYCNQYTGNTVELLPVLGM 1649  
QY 196 -P--G-----K-----N-----F----- 200  
Db 1650 SPEGYNLDHSTWHKYMGSHSIFIRYGNVMYPALDTPATHEALKLVKQFMQVINCRI 1709  
QY 201 -----DG-----R-----GM----- 205  
Db 1710 VVQAAPSSLAVAPSRTELSSQKMTEDGIVNLAVLVDMKEADIKARLPNAIEYIGKAIKE 1769  
QY 206 -----VI-----S-----N--ID----- 211  
Db 1770 YSPHRFEVNSYLPYTNLPDVIVSRYSMSSTLSKOYALCKKGNWEIDRQLEHAPALAHF 1829  
QY 212 TC-----R-H-----TD-----Y----- 216  
Db 1830 SGSQKAHLKHLKRAKDLNRYRELTKVFTDMAIKPMFKTFAKLGITKGISLFHYQYGVEL 1889  
QY 217 -----R-H----- 218  
Db 1890 WRNNIVNSVRRIHSLNFMFTYSSKKCRVFTVPTNIEDSLAGYPEPDNLSTETLVKYNIS 1949  
QY 219 -----KA-----M----- 221  
Db 1950 IVKVGTKKEAEAMTNLLTKAGYEVINMAEYHEWDCIARRRKEEQEAKLAISKLOAKKAT 2009  
QY 222 -----R-----I-----D-D-----D----- 226  
Db 2010 NKKVSPNRLISWNILEIPKEHRSYSSEYCGSHYASRSHDRMVDVDTPOYIMQEDINNY 2069  
QY 227 -----A--KA--S-----MR-----F-K 234  
Db 2070 TGRGLVSMWQLTPATIKATVVCVRNKIERKAIRKGAHVHDEWTAPYIASVLTSSKQFIK 2129  
QY 235 --KE-----D-----LKG----- 240  
Db 2130 VYTKERQKTLSDIGIYSKHLRLRLRELDITYAPLKLRLYDPELENAVLFLOSSYPKLRAL 2189  
QY 241 -----T-----D--K-----N--YW-----L 247  
Db 2190 LGKGLITEEYHFLSDVDVHCKTFNSRNLTLWQPDGSTWVIPOGDARVPRIVAEAKEKNL 2249  
QY 248 S-----D-----K--I-----PH----- 253  
Db 2250 TPGKIIEDVDTIEAQIRQBFDAERDTGGFVKFPRITKKLQEFPLGEEBQVPHVSPVE 2309  
QY 254 -----AFN-----Y--Y-----N----- 259  
Db 2310 LGAKPTAFNAFANAVGKTLAVQHASLKAPEPMTVTHERFYGNVAVMTDVPNRHRIE 2369

QY 260 -----G-----G-----K----- 262  
Db 2370 VIHVIRESGRTSEYVNNLLAQPTPFVKVDFWLTETQAKFVRRLNNTQGVISVYNKLEBDG 2429  
QY 263 I-----T-----V-----E----- 266  
Db 2430 IPVKVAVKATEPTNKEKLEAASTKLKANGAVATDAYEBHTDLNENEETIIVAVVNDKVVDP 2489  
QY 267 -----K-Y-----DG----- 271  
Db 2490 VQNLHRHIROSANLKYKFTKFLERLSTVIDKRRHSVEDLMKFMKGDPLIADDSGSI 2549  
QY 272 -----R-D----- 273  
Db 2550 FKRLSKONPEHLKDVMDVDFVDCSHKKIEOCVGMVVMKENLVQNRKRCDSHGLHVAS 2609  
QY 274 --Y-----F-----D-----P-----H----- 278  
Db 2610 LOYIRNPSGNVTIICKVAPEDVFAVPEYDVTKMRVSAHYIITKLPLDAIRDHVNNGNPIST 2669  
QY 279 --G-----M-----H----- 281  
Db 2670 IEGGTEILNMVLSGNHPSRQVLVGGHYGNLTYSSELTRNYLDDDDSVIPDEPVIKKT 2729  
QY 282 -----TA----- 283  
Db 2730 ALNMEESFTATNPKAEPVKATDVKPIKTKMEILKDLWNQFLKAETSNEAITLADEIIAA 2789  
QY 284 -----G-----I----- 285  
Db 2790 KGKCKKSNASLGFSQDMVQKLIIDARANKPKEKAVQDKTVAPKTTVTRSNADVIRSY 2849  
QY 286 L-----A-G----- 288  
Db 2850 LNDKMSDYCKAHAHIDVKRAAKSAAALGLTPBECSDIKLKHKLKHTNAKVYRANK 2909  
QY 289 -----N-----DT----- 291  
Db 2910 AHDDIIIRNNAVGLSGLTIAKTLGVHPTVTLRLSLGIEPADTRRTFMENVLAPLPHV 2969  
QY 292 --EQ-----D----- 294  
Db 2970 ADMLSEQVGPKEIRQYVYRDLLLLEAYNNRHNHKEHAHERFLKRYVVEVMELELAPNSEE 3029  
QY 295 -----I-----K----- 296  
Db 3030 AYDTLHNAIAALTELSELMSKSSDKVYAPNSNRLAILDALADQIVTATGIGTFLGMNVPG 3089  
QY 297 -----N-----FN-----G----- 300  
Db 3090 ALTEVNRNSYKFEDEGPVFNEKNKVMKEAAAGFFDLDPNASELRISGPGTGKTF 3149  
QY 301 --ID-----GIAP-----N-----A-----Q----- 309  
Db 3150 MAHMIDEIMPRYHETCSLMGI--PALYNEVIMTATTNKAAEVLARATGRPTSTYHSFQGLT 3208  
QY 310 -----IF-----S-----Y--K----- 314  
Db 3209 VRNLLKTGEADLIPSKSFSIKKNKVFIDEASMDRKLKKFINEGTHQSIVYLGDKCOL 3268  
QY 315 --M-----Y--S-----D-----A-----GS-----GF-----A----- 324  
Db 3269 LPVKTSPPVNDSSIKSFMLTQQMRTDVPELHALHEQLRGIEGHGFLPIKAVPGIID 3328  
QY 325 --GDE-----T-----M--PH-----A 332  
Db 3329 WIKDDEMRKEVEGHFMTKTDRIVAYTNQVYNTFIRAMNGFHGEFCIEELVSNSA 3388  
QY 333 I-----E-----DS----- 336  
Db 3389 IQIGAGDRLSIBQEQIVDQDSATRKILIDSSGIELEVRDCTLDTGCGYGVFEGIPIPVDM 3448

QY 337 --I-K--H--N-----V----- 341  
Db 3449 DYFIRLQGVYAHQKWBHFYLTKETPELRALHASTVHKSGQSTVDTFIDATDLSGRO 3508  
QY 342 -D- -VS-----V-----SS----- 348  
Db 3509 PDVAVRLLYVAVSRARHRVVEYGSIAEKFGELMQLGKQIREMTNSSELVOELFLAELR 3568  
QY 349 -----G--F-----TG-----TG----- 354  
Db 3569 RLDKVLSIITRNGETFAVYNSTGFLYLGEFYSSRSQOPPSGHRLLALNHLDDMKREY 3628  
QY 355 -----LV-----G----- 357  
Db 3629 LEQAGKLLLEVHNVNQVFRLLVQCKTYQDVADALPECLIAMDKQHLKSLQRTPKAFT 3688  
QY 358 -----E-----K- 359  
Db 3689 LDNDAMALRHNRKSGRLSNTWREPILSFEVCMRYITFPNNEGYPALITTLRKDLYKA 3748  
QY 360 Y--W-----O-----A-----I----- 364  
Db 3749 YMEFWDVDKDSVMIELHTAPGKKKTAKEMOEFITQELVPAFEAAQTQYIICGDSEYK 3808  
QY 365 --R-----A-----LR-----K-----A----- 370  
Db 3809 QLTQKQKADANIGVYMDCAVGNQKVIYVPNRQIFYPDCKIKTKIALSMQALIDVSSGY 3868  
QY 371 --G--I-----P-----M-----V-VAT-----G-----N- 380  
Db 3869 VEPGKTIHVADYPLTPRAISKWLDKLEMDVPLAVDIETSLKHYYGIGITFCWNKH 3928  
QY 381 -----Y-----ATSA-----S----- 386  
Db 3929 EGIAFPVDYBPIEGATEAPYRQVHNMLVRSMLDFIKYLRQMYNIAFVYALYQL 3968  
QY 387 --S-----S-----WD--LV-----A-----N 395  
Db 3989 FMTLLDTEGILHGHGISILRWMDCTKLTLYLATNSCAKNKSLDKQAQVAGNNAQOEIN 4048  
QY 396 -----N-----H--L--K--T-----D-----TG- 404  
Db 4049 DITRIPLAELLEVLVDGLCTWYVEKHMDTLVNDQLDVYTNIFKPACEDIIQWQLTGM 4108  
QY 405 -----N-----V-----TR----- 408  
Db 4109 PINNRVLEVEEALTDYNNALKTADSKVIAKDFRLLNEEWEKONQILKKRVTLLADA 4168  
QY 409 -----TA--A-----H-----E--DA-I 416  
Db 4169 KEQFNPSGIOLQKLLFEFLGLPVULGLTASKLPAATGSGILSLKNHTODTSLLEIDLALI 4228  
QY 417 --AV-----A--SA-KN--Q-----TV----- 426  
Db 4229 DYKAVDKILGTAFIPALKVARQGPDMHYLFGNLNLGTVSGRLSSSEBNQLNPSGRYA 4288  
QY 427 -----E-----F-----D--K----- 430  
Db 4289 KMIKCFEAPGWIIFCGLDFASLEDRISALTTPKPNKLVYTDGSHSLRAYVFGEM 4348  
QY 431 -----VN-I-----G-----G-----E----- 436  
Db 4349 PDIEDVESVNSIOEKYKAYRODSKAPTFALTQCTYITLMKNGCFPEQAKRMBEERYHT 4408  
QY 437 --S-----F-----K--Y----- 440  
Db 4409 LYKFSDDVQAQLDOAAKQGVTVAFGLRVRTPLQVIRGTSKTPYBAEAGRTAGNAL 4468  
QY 441 -----RN----- 442  
Db 4469 GQSMCLLNRAQSEFMKRVKNSKRLDIRPSTHIDAQFLIRDDMDVVIYTNHLVKAV 4528  
QY 443 -----I-----GA----- 445

Db 4529 QWQDHPDIAHPDVHLDDQEGAVIVQNTMLLTBEKQVTFRDLGRAQSALFQNLNDRFGQV 4588  
QY 446 -----F-----FD----- 448  
Db 4589 DVRDIVFLNLSYLGMSQRPDQONLVNPNORMEOPRYOAECEDIIOALNMTFDSGLFKA 4648  
QY 449 -----K-----S--K--I-- 452  
Db 4649 LMRVAAARLNGKRGKGNNAVYDAEKMVHYANRIFIKESGGQKKEABEPVSALKKGEELHM 4708  
QY 453 -----T-----T--N-----E-----DGT----- 459  
Db 4709 SLSELFTLLTGEAAEGEWSLAGNHTVPRRLGVMPRIAVMSKAEQDVVADGTTIHN 4768  
QY 460 -----K-AP----- 462  
Db 4769 FPHEDWSDKIAFPARITAMRFVWMLHDEYDVVKKKYLSVTLTKPIKIVNKHVLD 4828  
QY 463 -----S-----KL----- 465  
Db 4829 SVESMDIMDVSTSMGTGLHDSIEKAWNNGHNALRKLGTPERVQVNVINPTREQIASN 4888  
QY 466 -----KF-----V-----Y--I--GKG-----Q-- 474  
Db 4889 PNLIPWIEQATKKIAGWTIGKFDIVTBGLDQDVNSTYTYTWKG--GRDEEHRLOQSM 4947  
QY 475 -----D-----O-----D----- 477  
Db 4948 YRWLHDDKITEDVIRINVIPTDMKKALATSGENYRKRVHLKDIRLSYEATEEMIKAKL 5007  
QY 478 -LI--G-----L-----D-----L--R----- 484  
Db 5008 ALIDKYWDABESIEPECTDEELMRSDVFKYFADPEKAKORGASTKNPFTLFEARKFMA 5067  
QY 485 --GK-----I-----A-----VM-----DR-----I-Y----- 494  
Db 5068 EKGSGTILHVEIFRPMIDLTGVSNNRAIEIVDLCKTQNTDGFRRVEAYVGLAKMAS 5127  
QY 495 -----TKD-----L-----K--N-----A--FKK--AM----- 506  
Db 5128 CMGATITTKRGSEIRPVNIVAVALATSGFGKHSVNIVEDGFMAGFKKGFMEDTWRALAE 5187  
QY 507 -----DK-----GA-----R-----A 512  
Db 5188 RLWKIANERSARNATDQKEEFDKVEAEYRTGAYPTFDSGTPRAVQOLRHKLLLAGCGA 5247  
QY 513 I-----M-----V--V-----NT-----V----- 519  
Db 5248 INLOIDEIGSNLANADVLVFLVELYDQKIKOKULTKNTAESIRSEVDGKTPANLLFG 5307  
QY 520 -----N-----Y-----Y--N--BDN----- 526  
Db 5308 TPSKILNGSQTELEFYDFLDTGYSRCLFAVGQADKRAVLSQSAEETIRNLIKODNNAV 5367  
QY 527 -----W----- 527  
Db 5368 NKWANHFRSLADASKFGMKQTVEDDVGIALIDYKIOCEKQAAALADHEIRKAELEHRYF 5427  
QY 528 ----- 529  
Db 5428 KALKLAQAFVVDNSINVEHENLKQALLVEESGAVOTILNREKAYVLAKYASVTE 5487  
QY 530 -----LP-----A--M-----GY----- 535  
Db 5488 VTHADLEALPFPYKSGNAARNEMMTLATWGYKOHIIIKKTYNGIEFFRGETLKETDTN 5547  
QY 536 -----E-----A-----D-----E----- 539  
Db 5548 EMTVSGENFAYDYIAEKVPDOLHLVLTQOAGYHMANHFRGRHAEENALAGFNMIVID 5607  
QY 540 --G-----TK--S-----Q-----V-- 545

Db 5608 VDGTCSLDVCHLMKEYRFMTYTTKRTDBENRFRLLIPNNYQLELDGAEYKPFMDNVMA 5667  
QY 546 ---F---SI---S---G---D---D---G---V--- 554  
Db 5668 WLPFKTDESANQRAKAKWESCAKGTYYYNDAPLLDVRDFIPRTSKNEAFQOGMKQVQNLQ 5727  
QY 555 KL--W-----N-MI-----NP-----D--- 563  
Db 5728 NLERWFAQRIATGRNNHMIKAYALALVDSIGFNEVKQRVHEFNKKLSNPLSADLSTV 5787  
QY 564 ---K---K---T---E---V---K---RN----- 571  
Db 5788 MITVAKLLFLVEDCQLKEETMSEEFHGEWMDOLILIVGFSAAKGSASLRNIENQERWV 5847  
QY 572 --N---K-----E-----D---F-KDKLE-QY 583  
Db 5848 YLNTAEAGKRLPFKNFNGVRITDPYQILEYFDQIEHRDDVDGIIIVDSLTFMMDMLETQY 5907  
QY 584 -----Y-----P-I-----D---ME-SF----- 591  
Db 5908 VLTAAQTQAWGEFAQFKILOEKVVYKFAKPVIFTAHVKDEVDERAMELKTFFVPVKGSL 5967  
QY 592 --N-----SN-----K-----P---NVG----- 599  
Db 5968 KNGIEAYFTVVSARIDLKELEKYSNGMLEITEDEQELGYKHVQFQTRPTKKTGVGKIR 6027  
QY 600 -----D---E---K-E-ID---F----- 606  
Db 6028 SPNGMFSKQTYMNDNAQKLLDLHAEYYSTGQAAKAEKATDNLGGFGAKESDIYLANLK 6087  
QY 607 ---K---F---APD-TD-----KE-LY-----KE----- 619  
Db 6088 VAYAGKAASGANFTQIIA-DLTDLDGHSAGEYREQLYITSGTEKGCKCTYEKNGKEYFLP 6146  
QY 620 ---DI---I-----VP---A---G-----S 627  
Db 6147 GYTVINDILVMTSGETIPEAVFEKVVNVYDFDEKKEVAKSVMPVNAIGKFAVALKS 6206  
QY 628 ---T-----S-----K----- 629  
Db 6207 EEDKQTKDGSNGYVSTGETRFTNTIEKVPHDLHLTVVEABELTERGKELTVEBAVFWDK 6266  
QY 630 W-----G---P----- 632  
Db 6267 WLEKNKGVTDRDKTTKGASGKAGQPKPGATNTGAGASAAKSLFEIMOIPVLGVDPSPFN 6326  
QY 633 ---R---I---DL-LL---KPD---V---S-----A--- 644  
Db 6327 WGLARGMLDLETGILSGLDLKLVLTKPDGTVQRQNSKDMQAAEDITGVIDWFKEAKVI 6386  
QY 645 ---P-G---KN-IKS-----T-LN---V---I-N-----G-K-S----- 660  
Db 6387 FVEVPVGSQSANGMKSYGVCVILGALTALGHEIIEVTPIENKVVALSGIKTASKDMVIRA 6446  
QY 661 -T--Y-----GY 664  
Db 6447 ATGFYPEANWLKDTKEFNENFIKGITMQVILTQPEIEAALTGY 6489